Visualizing Set Relations and Cardinalities Using Venn and Euler Diagrams

by

Luana Micallef

September 2013

(revised version approved February 2015)

A Dissertation Submitted to the University of Kent, UK
in the Subject of Computer Science in
Partial Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

© Luana Micallef, 2013
University of Kent, UK

Abstract

In medicine, genetics, criminology and various other areas, Venn and Euler diagrams are used to visualize data set relations and their cardinalities. The data sets are represented by closed curves and the data set relationships are depicted by the overlaps between these curves. Both the sets and their intersections are easily visible as the closed curves are preattentively processed and form common regions that have a strong perceptual grouping effect. Besides set relations such as intersection, containment and disjointness, the cardinality of the sets and their intersections can also be depicted in the same diagram (referred to as area-proportional) through the size of the curves and their overlaps. Size is a preattentive feature and so similarities, differences and trends are easily identified. Thus, such diagrams facilitate data analysis and reasoning about the sets. However, drawing these diagrams manually is difficult, often impossible, and current automatic drawing methods do not always produce appropriate diagrams.

This dissertation presents novel automatic drawing methods for different types of Euler diagrams and a user study of how such diagrams can help probabilistic judgement. The main drawing algorithms are: eulerForce, which uses a force-directed approach to lay out Euler diagrams; eulerAPE, which draws area-proportional Venn diagrams with ellipses. The user study evaluated the effectiveness of area-proportional Euler diagrams, glyph representations, Euler diagrams with glyphs and text+visualization formats for Bayesian reasoning, and a method eulerGlyphs was devised to automatically and accurately draw the assessed visualizations for any Bayesian problem. Additionally, analytic algorithms that instantaneously compute the overlapping areas of three general intersecting ellipses are provided, together with an evaluation of the effectiveness of ellipses in drawing accurate area-proportional Venn diagrams for 3-set data and the characteristics of the data that can be depicted accurately with ellipses.

Acknowledgements

Finally I am done! The long and challenging journey has come to an end. This would not have been possible without all the people who have contributed to my work and my sanity throughout these years.

Many many many thanks to my supervisor, Dr Peter Rodgers, for his continuous support, advice and guidance. He introduced me to the field of Information Visualization and through him, I had the opportunity to connect and collaborate with other experts, including Angela Morelli and Dr Gem Stapleton. I would like to thank Angela Morelli for her feedback on the design of the diagrams and for sharing her enthusiasm for information design. I am grateful to Dr Gem Stapleton for her time and advice on the theoretic aspects of this work.

A special big thank you goes to Prof Jean-Daniel Fekete and Dr Pierre Dragicevic for giving me the opportunity to intern with the AVIZ team at INRIA Saclay, France and to work with them on the application side of my PhD research. It was a great enriching and inspirational experience, and I would like to thank all of AVIZ for being so amazing. I am grateful to INRIA for funding my internship and to Prof Jean-Daniel Fekete and Dr Pierre Dragicevic for their continuous valuable advice.

I must thank the School of Computing at the University of Kent, UK for funding this research, and all the staff and officemates for the great working environment. Also, The Allan & Nesta Ferguson Charitable Trust for partially funding the final year of this research.

Finally, I would like to thank all of my family for being there for me at all times, particularly my number one supporter, my mum.

To Odette, Aiden and Nicholas

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

This chapter outlines the motivation, objectives, contributions and structure of this dissertation.

1.1 Motivation and Objectives

Advances in technology have led to the vast amount of data that is routinely generated. Emails, digital photos, CCTV footage, tracking data from smart devices are a few examples of the types of the data that anyone can generate on a daily basis. Through this data, links between communities can be identified and crimes or terrorist attacks can be predicted and disrupted. Historical stock market data is studied and compared to identify market trends and to aid investors make informed decisions. Data from scientific experiments is generated to analyse relationships between groups of genes and find cures to illnesses. So often, data sets are compared, and relationships between data elements and sets are studied.

Various visual representations of relationships between data elements and data sets have been proposed to facilitate analysis. However, similarities and differences are not always easily visible. Consider this example. The geometric shapes in Figure 1.1A are all quadrilaterals, as they all have four sides, four vertices and interior angles that add up to 2π. There are different types of quadrilaterals. Each type is distinguished by specific properties (e.g., a rectangle has all four angles right angles and opposite sides parallel and equal in length), and some types share a few properties (e.g., a rectangle and a parallelogram both have opposite sides parallel and equal in length). Thus, the quadrilaterals in Figure 1.1A are of one or more types, and various quadrilaterals could be of a certain type.
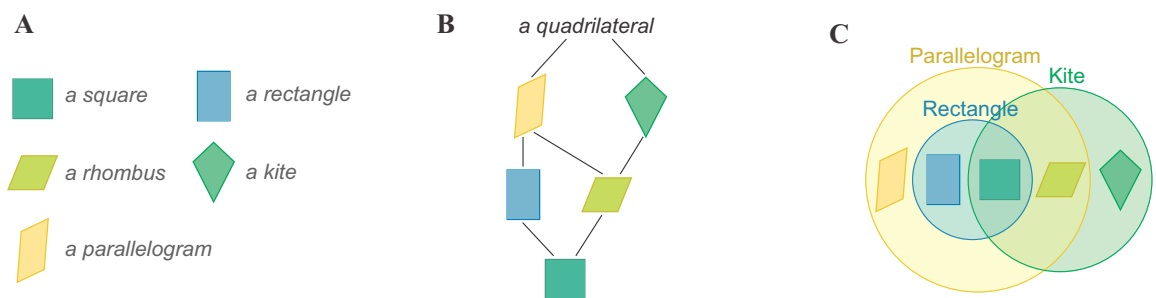
Figure 1.1: *Visualizing relationships between different quadrilaterals (left, A) using a node-link diagram (centre, B) and an Euler diagram (right, C).*

The relationships between the quadrilaterals, based on their properties and types, can be represented by a hierarchical tree and visualized using a node-link diagram as in Figure 1.1B. Instead of using links, closed curves could be used, such that the quadrilaterals of a specific type are enclosed within one region, as illustrated by the Euler diagram in Figure 1.1C.

From Figure 1.1C, we immediately note that, for instance: there are kites that are also parallelograms and rectangles; a rhombus is both a parallelogram and a kite; and a square is a rectangle, a parallelogram and a kite as well as a rhombus, since the rhombus is a parallelogram and a kite. The same information can be derived from Figure 1.1B, but the paths have to be traversed and information is discovered level by level. For instance, the square has two links indicating that it is a rectangle and a rhombus. Moving up a level, we then get to know that a square is also a parallelogram and a kite, since the rectangle and the rhombus are parallelograms, and the rhombus is a kite. This means that the shapes have to be considered one at a time for their types to be identified and so, similarities and differences between the quadrilateral types are not so easily noticeable. The levels in the hierarchy could facilitate reasoning about containment, but intersections and exclusions are not easily noticed.

The similarities and differences between the quadrilaterals types are more easily noticeable in Figure 1.1C than in Figure 1.1B. This is because in Figure 1.1C, shapes with certain properties are bounded by closed curves, forming common regions [Palmer, 1992] that have a perceptual grouping effect that is stronger than and independent from any other grouping principles, such as the Gestalt powerful proximity and similarity [Koffka, 1935]. Closed curves are also processed preattentively, so that the curves and their intersections are distinguishable and seem to pop out as discrete objects without requiring focused attention [Enns, 1986; Treisman and Souther, 1985]. Moreover, a single diagram can depict different types of relationships (intersection, containment, exclusion). All of these features facilitate analysis and reasoning about relationships between concepts. Thus the reason why Euler [1768a] successfully explained categorical propositions to laymen using the diagrams, and why they are used in various areas such as medicine, finance and criminology to facilitate data analysis.

Before drawing an Euler diagram, the data sets and the intersections that need to be depicted are first identified. If all the possible intersections between the identified data sets are required, then the Euler diagram is a Venn diagram [Venn, 1880]. Drawing a closed curve for each identified data set and overlapping the curves to depict only the required set intersections is difficult, particularly when certain aesthetic features should be satisfied, so that the generated diagram is comprehensible and usable.

In this dissertation, we devised novel automatic drawing algorithms for different types of Euler diagrams and we conducted a user study to assess the effectiveness of Euler diagrams for probabilistic judgement. To develop algorithms that draw accurate Venn and Euler diagrams with aesthetically desirable features, we explored and adopted various techniques. Firstly, we investigated force-directed graph drawing techniques and we devised a force-directed method, eulerForce, to lay out Euler diagrams. Secondly, we examined the effectiveness of ellipses in drawing Venn diagrams that accurately depict the cardinality of 3-set data through the size of the curves and their overlaps, and we devised a method, eulerAPE, to automatically draw these diagrams using ellipses. Finally, we assessed the effect of area-proportional Euler diagrams, glyph representations, Euler diagrams with glyphs and text+visualization formats on Bayesian reasoning, and we devised a method, eulerGlyphs, to automatically draw such visualizatons for Bayesian problems. We further discuss the motivation and objectives of this work in the next sections.

1.1.1 Laying Out Euler Diagrams Using a Force-Directed Approach

Circles are regular and have good continuity and so, Euler diagrams with curves depicted as circles are easier to understand than those with non-smooth curves [Benoy and Rodgers, 2007]. However, an Euler diagram cannot be drawn with circles for all possible data, unless unwanted set intersections are depicted [Stapleton et al., 2012]. Drawing methods that automatically generate Euler diagrams with only the required set intersections have been proposed (e.g., [Rodgers et al., 2008b]). However, the curves of these generated diagrams are often non-smooth and not easily distinguishable. Figure 1.2A is an example. It accurately depicts the required data, but it is difficult to comprehend and rather unusable.
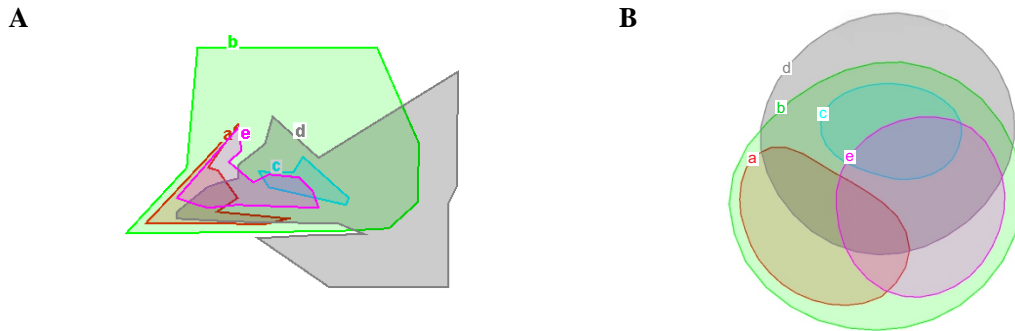
Figure 1.2: *An Euler diagram generated by an automatic drawing method [Rodgers et al., 2008b] that only depicts the required set intersections (left, A), and its improved layout generated by our force-directed method,* euler*Force (right, B).*

Layout methods could be devised to take these diagrams and improve their layout, so that the required data is still depicted accurately, but the diagram is easier to comprehend. Two such methods have been proposed: one uses graph transformations [Rodgers et al., 2010b] and another uses multi-criteria optimization [Flower et al., 2003b]. However, only the latter has been implemented and even so, the method is complex, computationally expensive and impractical when a number of aesthetic criteria have to be satisfied and when the diagrams have various curves and intersections. In contrast, force-directed techniques have been successfully used to lay out node-link diagrams of graphs with various vertices and edges in relatively fast time and with various desirable aesthetic features [Di Battista et al., 1999a; Eades, 1984; Gibson et al., 2012; Kobourov, 2012].

Euler diagrams are a collection of overlapping curves. If a closed curve is represented as a polygon and the basic attractive and repulsive forces of such force-directed methods are applied to the vertices and edges, the curve would be laid out as a regular circle-like polygon. If all the curves in the diagram are laid out in this way, the curves would be easily distinguishable and the diagram would be easier to comprehend. So such an approach could improve the layout of the diagrams in relatively fast time. Since force-directed algorithms are often efficient and not complex to implement, different forces can be used to handle different aesthetic features. However, the challenge is to develop a force model that improves the layout and the curve aesthetics of the diagram, and maintains the diagram's depicted set intersections. Prior to our work, force-directed techniques have not been used to lay out Euler diagrams.

We developed the force-directed method euler*Force* to lay out Euler diagrams, so that Figure 1.2A is converted to Figure 1.2B in relatively fast time.

1.1.2 Drawing Area-Proportional Venn Diagrams Using Ellipses

Venn and Euler diagrams can also depict set intersection cardinalities or any other quantitative data assigned to the sets and their intersections, in addition to depicting relationships between data sets. Such diagrams are said to be area-proportional as the total area of the regions corresponding to a specific set intersection is directly proportional to the cardinality of or the quantity assigned to that intersection. Size is a visual variable [Bertin, 1983] and a preattentive feature [Treisman, 1985] that is easily noticeable due to its pop-out effect [Treisman and Gelade, 1980], which is classified third in strength just after colour and orientation [Ware, 2012]. This further aids data analysis and the identification of differences and similarities between data sets.

Area-proportional Venn diagrams with three curves are extensively used in diverse areas, particularly for medical and scientific disciplines. An informal study identified various area-proportional Venn diagrams (72 diagrams) in 2009 in the world's most cited journals [Wilkinson, 2012]. Looking into these diagrams, almost all have two or three curves and are drawn using circles. Most of those with three circles are misleading and inconsistent with the data. Figure 1.3A is an example obtained from a BMC Medicine journal paper [Lenz and Fornoni, 2006] illustrating the results (the numeric labels) from a survey that assessed whether medical trainees can manage chronic kidney disease. This diagram was meant to raise awareness that better guidelines on the screening and therapy of the disease should be provided. Yet the regions with labels 1% and 25% are respectively bigger than those with 3% and 29%.
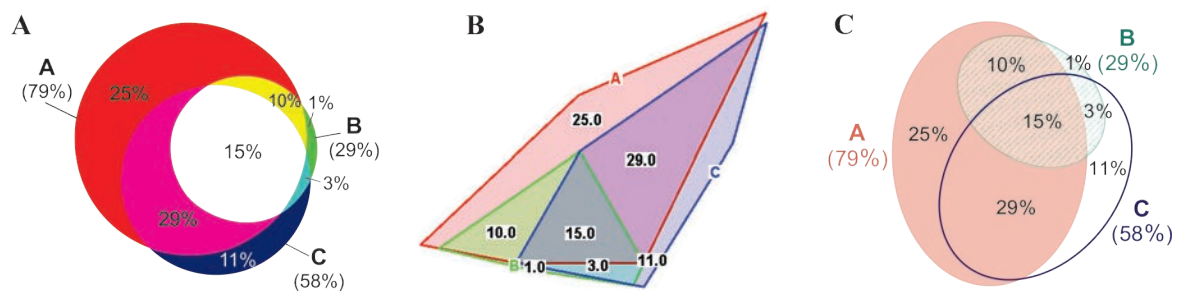
Figure 1.3: *Area-proportional Venn diagrams generated by different drawing methods for the same quantitative data (indicated by the numeric labels) using circles, polygons, and ellipses.* (A) A redrawing of Figure 5 (bottom) in the medical journal article [Lenz and Fornoni, 2006] from where the data was obtained. The authors generated this diagram using the first drawing method that used circles [Chow and Rodgers, 2005]. (B) Generated by Convex Venn-3 using 4-sided and 5-sided polygons [Rodgers et al., 2010a]. (C) Generated by our drawing method eulerAPE using ellipses.

Such area-proportional Venn diagrams cannot be drawn analytically using a specific curve shape and so, numerical methods or heuristics are required [Chow, 2007]. Circles can draw accurate area-proportional Venn diagrams for any data with two sets [Chow and Ruskey, 2004], but not three [Chow, 2007] due to their limited degrees of freedom (i.e., a centre and a radius). Polygons can draw accurate area-proportional Venn diagrams for any 3-set data, but their non-smoothness impedes comprehension [Benoy and Rodgers, 2007]. Yet current automatic drawing techniques represent the curves as either circles or polygons. The authors of the BMC Medicine article generated Figure 1.3A using the first drawing method that used circles [Chow and Rodgers, 2005]. For the same data as that in Figure 1.3A, we generated Figure 1.3B using a drawing method that uses 4-sided and 5-sided polygons [Rodgers et al., 2010a]. Though Figure 1.3B accurately depicts the data, the curves are not easily identified without

focused attention [Treisman and Souther, 1985]. So despite the inaccurate diagrams, circles are often preferred over polygons due to their good continuity [Koffka, 1935], which enables the curves to be easily identified [Field et al., 1993; Hess and Field, 1999] and to pop out as discrete and complete objects [Ware, 2012], so that the relationships between the sets are easily visible.

Ellipses, on the other hand, have more degrees of freedom than circles and are similarly smooth. So diagrams drawn with ellipses are more likely to depict the required 3-set data accurately than those with circles. Thus, automatic drawing methods that use ellipses should be developed and the effectiveness of ellipses in accurately depicting most 3-set data should be investigated. The need to use such curves was previously suggested [Chow and Ruskey, 2004; Chow, 2007; Wilkinson, 2012], but was not adopted prior to the work in this dissertation due to difficulties in calculating the area of the regions of overlapping ellipses and in adjusting the various properties of the ellipses [Chow and Ruskey, 2004].

We developed analytic algorithms that instantaneously compute the area of the regions of three general intersecting ellipses and we devised the drawing method eulerAPE to automatically draw area-proportional Venn diagrams with three curves using ellipses. As shown in Figure 1.3C, an accurate area-proportional Venn diagram with smooth curves can be drawn with ellipses (using eulerAPE) for the same data as that in Figure 1.3A, which is inaccurate, and that in Figure 1.3B, which is difficult to comprehend. The design choices of the various components that make up eulerAPE (e.g., the heuristic algorithm, cost function, starting diagram generator) and the final drawing method were thoroughly evaluated to ensure its effectiveness in drawing a good diagram when one is known to exist for the required data. With such a drawing method, we aspired to evaluate the effectiveness of ellipses in drawing accurate diagrams with smooth curves for random and real world 3-set data (in contrast to circles and polygons used by previous methods), and to identify characteristics of data that can or cannot be depicted accurately with ellipses.We focused our work on Venn diagrams with three curves, so that an extensive evaluation of both the drawing method and the effectiveness of ellipses could be carried out, paving the way for future extensions to handle Euler diagrams with any number of curves.

Design features, such as colour and outline, may hinder the identification of the sets and their intersections. Figure 1.3A uses circles, but identifying the curves in which the regions are located is still difficult as the colours of the regions located within the same curve are completely unrelated. So based on theories of human perception and cognition, we aspired to devise diagram design features that aid reasoning about the relationships between the sets, by ensuring that separate layers, one per curve, are perceived and that the curves in which regions are located are easily identified, as in Figure 1.3C.

1.1.3 Assessing Euler Diagrams with Glyphs for Bayesian Reasoning

Area-proportional Euler diagrams could be particularly problematic for Bayesian reasoning, as biases elicited by area judgement [Cleveland and McGill, 1984] would add on to the cognitive judgemental errors and biases that affect Bayesian reasoning [Bar-Hillel, 1980; Barbey and Sloman, 2007a, b]. Studies in Bayesian reasoning illustrate that Euler diagrams clarify the relationships between the quantities and the sets of Bayesian problems [Sloman et al., 2003], while frequency grids as in Figure 1.4A facilitate logical reasoning by depicting the quantities as discrete objects [Brase, 2009; Cosmides and Tooby, 1996]. So combining the two into one hybrid diagram as in Figure 1.4B could aid, as the curves of the Euler diagram only depict the sets and their intersections, while equally-sized glyphs (proportional in number to the required quantities) depict the quantitative data. Such a diagram

could thus be a good alternative to an area-proportional one. The use of 'equal-sized points' proportional in number to the required quantity was also suggested by Bertin [1983] as an alternative to area.

However, it is still unclear which is the most effective visualization to aid Bayesian reasoning and the few studies carried out in psychology had a number of limitations. For instance, most studies were carried out over a small, non-diverse population, typically highly educated students (e.g., [Cole and Davidson, 1989; Cosmides and Tooby, 1996]). In each study, one or very few visualizations were tested, often over one problem (e.g., [Brase, 2009; Cosmides and Tooby, 1996]). In various cases, the visualizations were inconsistent, depicting the data incorrectly (e.g., [Brase, 2009; Sloman et al., 2003]).
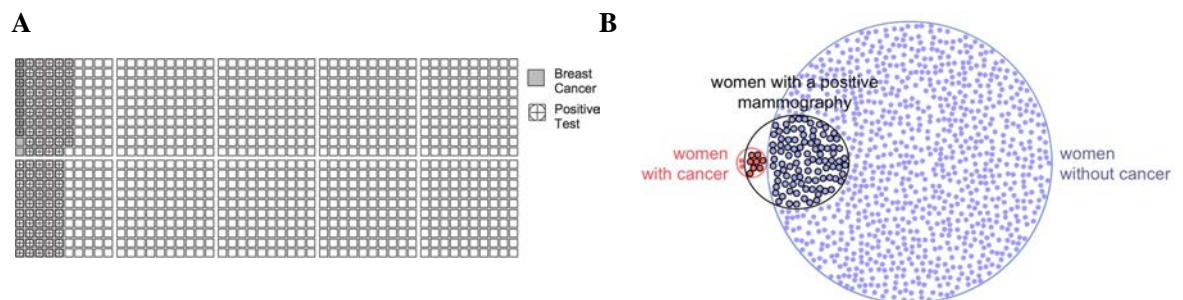
Figure 1.4: *A frequency grid and an Euler diagram with glyphs for the classic mammography Bayesian problem [Eddy, 1982].* (A) A frequency grid. *Source:* [Sedlmeier and Gigerenzer, 2001]–Figure 5 (B) One of the Euler diagrams with glyphs assessed in our study and generated with our method, euler*Glyphs*.

Both laymen and professionals have difficulty in probabilistic judgement and Bayesian reasoning and are unaware of their wrong judgements. However, based on such reasoning, important decisions that could have severe consequences (e.g., overdiagnosis of diseases, not predicting criminal attacks) have to be made. Euler diagrams with glyphs seem to have the right features to facilitate reasoning of both the set relations and the quantitative data. Thus, we devised euler*Glyphs*, the first method that draws two different types of area-proportional Euler diagrams, a glyph representation and three different types of Euler diagrams with glyphs (Figure 1.4B represents one type) for probabilistic judgement, and we conducted a study through crowdsourcing on three Bayesian problems to assess whether any of the six visualizations generated by euler*Glyphs* can facilitate reasoning and decision-making. Through this study, we aspired to identify the effectiveness of the different visualizations on a large, diverse group of participants and over various Bayesian problems and to determine whether depicting quantitative data using discrete objects is more effective than using area. We also aspired to investigate novel ways how visualizations could be holistically combined with textual information. To our knowledge, this is the only computer science study on Bayesian reasoning since 1989 [Cole and Davidson, 1989; Cole, 1989].

1.2 Contributions

The major contributions of this dissertation are the development (and evaluation) of novel drawing algorithms for different types of Euler diagrams, and a user study assessing the effectiveness of Euler diagrams for Bayesian reasoning. More specifically:

1. the development of the method, eulerForce, that uses a force-directed approach and a novel force model to automatically lay out Euler diagrams (Chapter 4);

2. the development of the method, eulerAPE, that uses ellipses to automatically draw accurate area-proportional Venn diagrams with three smooth curves (Chapter 5), including also:

   a. the development of analytic geometry algorithms that instantaneously compute the area of the regions of three general intersecting ellipses;

   b. the development of (i) a novel cost function to direct the optimization to the required diagram, (ii) an algorithm to generate a rational starting diagram for the optimization, (iii) an algorithm to adjust the ellipses' properties during the optimization process;

   c. the evaluation of the effectiveness of ellipses in drawing accurate area-proportional Venn diagrams with smooth curves, in contrast to circles and polygons, for random 3-set data and for real world data, and the identification of characteristics of the quantitative 3-set data that can be depicted accurately using ellipses;

   d. diagram design features that could facilitate comprehension of the diagram and the depicted set relations and their set intersection cardinalities;

3. a user study assessing the effect of area-proportional Euler diagrams, glyph representations, Euler diagrams with glyphs and text+visualization formats on Bayesian reasoning, (Chapter 6) including also:

   a. recommendations on how visualizations should be combined with textual information;

   b. the development of the method, eulerGlyphs, that automatically draws the assessed diagram types for any Bayesian problem.

The devised drawing algorithms generate accurate diagrams with aesthetically desirable features that could facilitate analysis of data set relations and their intersection cardinalities, assisting users in knowledge discovery and decision-making. Our evaluation of these drawing algorithms can aid in the development of other improved drawing methods in the future. Finally, our user study on Bayesian reasoning will guide future research on how quantitative data should be depicted with Euler diagrams, how to facilitate probabilistic judgement, and text+visualization formats that should be considered next.

To promote these contributions and ensure accessibility, the work in this dissertation has been presented and published in journals and conference proceedings. Software implementations of our novel algorithms, as well as the experimental stimuli and data of our user study are also available online.

1.3 Publications

The contributions of this dissertation (Section 1.2) have appeared in the following publications.

1. Refereed journal article, eulerForce: *Force-directed Layout for Euler Diagrams*, in the Journal of Visual Languages and Computing (JVLC) [Micallef and Rodgers, 2014b]. This paper was accepted at the 2014 International Workshop on Visual Languages and Computing (VLC) as part of the 2014 International Conference on Distributed Multimedia Systems (DMS). Papers accepted at DMS (including workshops) are presented at the conference and appear in a special issue of the JVLC after an additional round of reviews. This work appears in Chapter 4, which discusses the first contribution of this dissertation (contribution 1).

2. Refereed extended abstract and poster, *Force-Directed Layout for Euler Diagrams*, at the 2009 IEEE Information Visualization (IEEE InfoVis), one of the main conferences of IEEE VIS (formerly VisWeek) [Micallef and Rodgers, 2009]. This work appears in Chapter 4, which discusses the first contribution of this dissertation (contribution 1).

3. Refereed journal article, eulerAPE: *Drawing Area-Proportional 3-Venn Diagrams Using Ellipses*, in the open-access PLoS ONE [Micallef and Rodgers, 2014a]. This work appears in Chapter 5, which discusses the second contribution of this dissertation (contribution 2).

4. Refereed workshop paper, *Computing the Region Areas of Euler Diagrams Drawn with Three Ellipses*, in the Proceedings of the 2014 International Workshop on Euler Diagrams as part of the 2014 International Conference on the Diagrammatic Representation and Inference (Diagrams) [Micallef and Rodgers, 2014c]. This work appears in Section 5.4 of Chapter 5, which discusses contribution 2a.

5. Refereed conference poster, *Drawing Area-Proportional 3-Venn Diagrams Using Ellipses*, at the 2012 Annual Grace Hopper Celebration for Women in Computing (GHC), Poster Session and ACM Student Research Competition (SRC) [Micallef and Rodgers, 2012]. GHC 2012 had an overall acceptance rate of 13.32%. This work appears in Chapter 5, which discusses the second contribution of this dissertation (contribution 2).

6. Refereed journal article, *Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing*, in the IEEE Transactions on Visualization and Computer Graphics (IEEE TVCG) 18(12) [Micallef et al., 2012]. This paper was accepted at the 2012 IEEE Information Visualization (IEEE InfoVis) conference and received one of three Best Paper Honourable Mention Awards. IEEE InfoVis 2012 had an acceptance rate of 24.72%. All papers accepted at IEEE InfoVis are presented at the IEEE VIS (formerly VisWeek) and appear in a special issue of the IEEE TVCG (volume 18, issue 12, for 2012). This work appears in Chapter 6, which discusses the third contribution of this dissertation (contribution 3).

7. Refereed conference paper, *Visualizations with Venn and Euler Diagrams*, in the Proceedings of the 2012 Annual Grace Hopper Celebration for Women in Computing (GHC), PhD Forum [Micallef, 2012]. GHC 2012 had an overall acceptance rate of 13.32%. This paper summarizes the contributions of this dissertation that are discussed in Chapter 4, Chapter 5 and Chapter 6.

8. Refereed conference paper, *Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges*, in the Proceedings of the 2014 Annual Eurographics Conference on Visualization (EuroVis), State of the Art Reports (STARs) [Bilal et al., 2014]. This paper was invited for submission to the Computer Graphics Forum (CGF) journal. As the second author of the paper, I wrote the sections related to Euler diagrams. These sections summarize the literature review in Chapter 3.

The software implementations of our algorithms and the other resources available online in relation to our work (e.g., experimental stimuli and data; related videos) are listed in the following section.

1.4 Software Implementations and Online Resources

The novel drawing algorithms discussed in this dissertation have been implemented (in Java) and the software implementations are available online as freeware. The source code is also available under GNU General Public License version 3 (GPLv3).

The software implementations are available at:

euler**Force**   `http://www.eulerdiagrams.org/eulerForce`

euler**APE**   `http://www.eulerdiagrams.org/eulerAPE`

euler**Glyphs**   `http://www.eulerdiagrams.org/eulerGlyphs`

Resources corresponding to the third contribution of this dissertation (contribution 3, Section 1.2; publication 3, Section 1.3) are available online for public reference and possible use in future studies:

1. Experimental stimuli and experimental data, in relation to our user study assessing the effect of visualizations and text+visualization formats on Bayesian reasoning [Micallef et al., 2012], are available at
    `http://www.aviz.fr/Research/Bayes`

2. A video with a visual explanation of a classic Bayesian problem using Euler diagrams with glyphs is available at
    `http://www.youtube.com/watch?v=D8VZqxcu0I0`

1.5 Overview of Chapters

After this introduction chapter, two chapters provide a literature review on areas related to this work, followed by a chapter on each of the three contributions in Section 1.2 and a conclusion chapter.

Chapter 2

A literature review on information visualization, visual representations of data set relations, graph drawing techniques particularly those using a force-directed approach, visualizations for Bayesian reasoning, and effective methods for combining visualization and text.

Chapter 3

A literature review on Venn and Euler diagrams, including: the historical background, application areas, diagram properties, aesthetic and perceptual factors, automatic drawing techniques, area-proportional Venn and Euler diagrams, Euler diagrams with glyphs, and Euler diagrams for Bayesian reasoning.

Chapter 4

Discusses the first contribution of this dissertation, that is the drawing method euler*Force* that uses a force-directed approach (similar to that used in graph drawing) to automatically lay out easy to comprehend Euler diagrams. The force model and the algorithm are discussed, together with our experimental results, evaluation and related future work.

Chapter 5

Discusses the second contribution of this dissertation, that is the drawing method euler*APE* that uses ellipses to automatically draw accurate area-proportional Venn diagrams with three smooth curves. The drawing method and our analytic geometry methods that instantaneously compute the area of the regions of three general intersecting ellipses are discussed, together with our evaluation of the effectiveness of ellipses in drawing such diagrams, characteristics of the quantitative data that can be depicted accurately using ellipses, diagram design features that could facilitate comprehension and related future work.

Chapter 6

Discusses the third contribution of this dissertation, that is our user study on the effectiveness of area-proportional Euler diagrams, glyph representations, Euler diagrams with glyphs and text+visualization formats for Bayesian reasoning. Guidelines on how visualizations should be combined with textual information to facilitate reasoning are provided. The drawing method euler*Glyphs* that automatically draws the diagram types assessed in our user study for any Bayesian problem is also discussed.

Chapter 7

Summarizes the objectives, contributions and findings of this dissertation and highlights important future work.

Chapter 2

General Background

This chapter provides a literature review on areas relevant to this research. This will help the reader understand the novel research work discussed in this dissertation. It excludes Venn and Euler diagrams, since these are discussed in detail in the next chapter.

2.1 Introduction

Graphics are used extensively everywhere and everyday. Traffic and public signs are simple examples of such graphics. Their effectiveness is essentially due to their simplicity in quickly and universally conveying messages and information. In some cases, such as metro maps, a graphic encodes a large amount of information, which could facilitate reasoning and enable faster and more accurate decision-making. It is for this reason why Arthur Brisbane stated, "Use a picture. It's worth a thousand words" [Syracuse Post Standard, 1911]. However, "a picture worth a thousand words must first be a good picture" [Bowman, 1967]. A visualization is good if the information obtained from it, when it is visually decoded, is the same as the original information that was graphically encoded [Cleveland, 1994]. Various visual representations of information have been and are still being proposed to facilitate data analysis and decision-making for diverse tasks and applications. However, a good visualization is often not enough. Text is typically used to convey specific details and so, the visualization and the textual information should be holistically combined to ensure that the link between the two is easily noticeable, such that the two are seen as one and that one is seen as incomplete without the other [Tufte, 1983].

We start this review with an introduction to information visualization (Section 2.2), followed by an overview of different visual representations of data set relations (Section 2.3) and techniques used in graph drawing to visualize relationships between data items using a node-link diagram (Section 2.4). The effectiveness of visualizations has been studied in various application areas, one of which is Bayesian reasoning. We discuss visualizations for Bayesian reasoning (Section 2.5) and a few techniques on how visualizations should be combined with text to facilitate comprehension and reasoning (Section 2.6). We then conclude with a summary of this chapter (Section 2.7).

2.2 Information Visualization

Visual representations of data have been used for millennia be it for presentation, communication and storytelling, or as a means to understand, analyse and reason about data, make decisions and answer questions [Friendly, 2008; Mackinlay, 2009; Spence, 2007; Tufte, 2001].

Cave paintings during the Ice Ages (after 25,000 BC) were among the earliest visualizations, followed by maps like those on Babylonian clay tablets and the wall painting of Çatalhöyük (Figure 2.1) [Friendly and Denis, 2001]. Early writing inscriptions in 3100 BC such as the Mesopotamian cuneiform and the Egyptian hieroglyphs respectively emerged from pictography and the civilizations' pre-literate artistic customs [Robinson, 2007]. In geometry, visualizations have been used since its roots in ancient Mesopotamia and Egypt in second millennium BC for astronomy, architecture and other purposes. In around 300 BC, Euclid included various diagrams in his treatise on geometry, *Elements* [Heath, 1956].
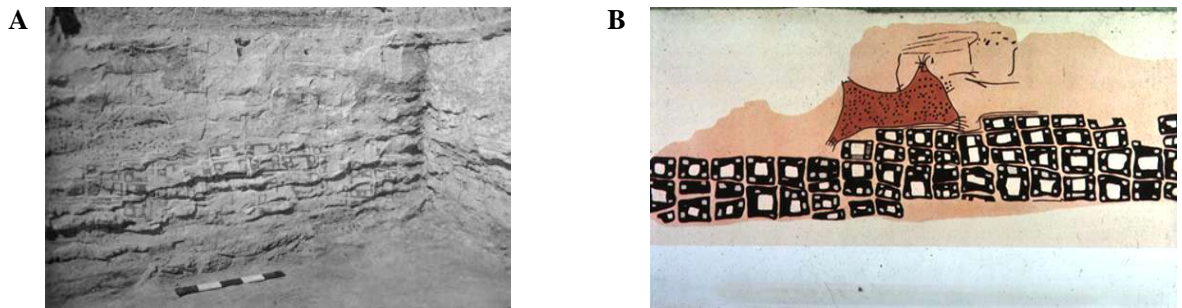
Figure 2.1: *An early map (7500 BC) depicting the city of Çatalhöyük, Antolia, Turkey. Source*: [Friendly and Denis, 2001] (A) The original wall painting. (B) A recreation of A showing the city's structure.

Visualizations have long been used to report findings and observations in astronomy (e.g., [Galilei, 1613], [Scheiner, 1626]) and to present statistical data (e.g., [van Langren, 1644, p. 3]—the first known statistical graph). Playfair (1759-1823), "the founder of graphic methods of statistics" [Fitzpatrick, 1960, p. 38], invented and used the bar chart and the time-series line chart to, for instance, summarize England's balance of trade [Playfair, 1786], and the circle chart and the pie chart to depict quantities, such as the cities' population (Figure 2.2A) and the countries' land mass (Figure 2.2B), as area [Playfair, 1801]. Using these graphics, he aimed to better communicate statistical data by exploiting the human psychological capabilities [Spence, 2006].

Playfair's graphics have inspired other classics in information visualization. Nightingale [1858] used the area of the radial segments in her iconic 'rose' diagram to report the number of deaths during the Crimean war (Figure 2.3). Minard [1869] used a flow map with two lines of varying thickness to show the loss of men in Napoleon's army as they moved from Poland to invade Russia and as they later retreated, between 1812 and 1813 (discussed in Section 2.6, Figure C.8). Dr John Snow marked deaths from cholera for September 1854 with dots and water pumps with crosses on a map of a London district and from the concentration of dots in a specific area, he managed to identify the contaminated pump (visualization available in Tufte's book [1983, p. 24]). Thus, with a simple visualization, Snow managed to answer his question and save lives. A more recent classic is Harry Beck's 1931 London Underground map [Garland, 1994], which, with a schematic design that is topologically exact as the true route but geographically inaccurate (as underground the exact geographic location does not matter), the map is easily processed and remembered in full detail, as a mental image of the map is easily constructed.

A

B
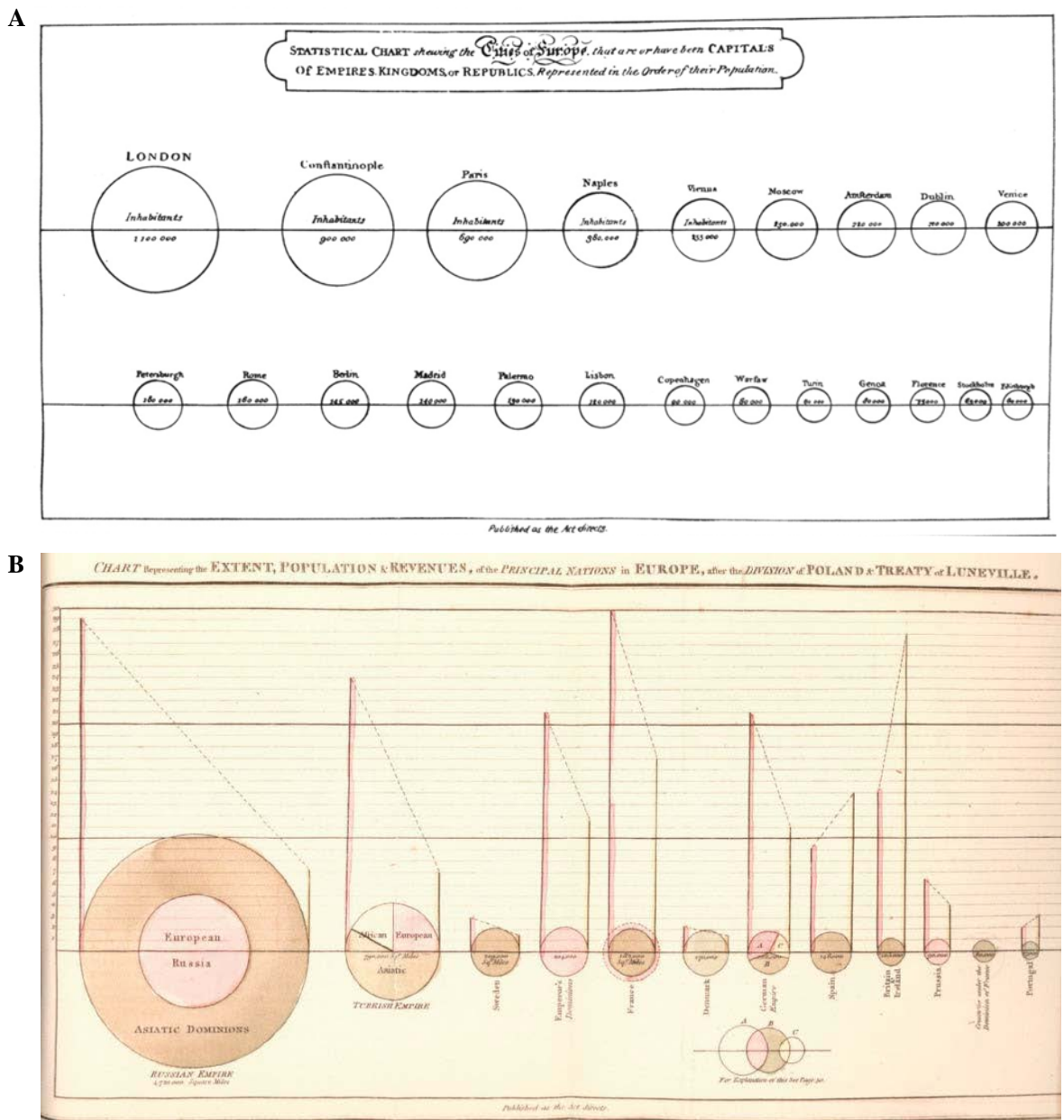
Figure 2.2: *First applications of area to represent quantities—William Playfair.* These diagrams are from the book 'Statistical Breviary' [Playfair, 1801], which Tufte describes as Playfair's "most theoretical book about graphics" [Tufte, 1983, p. 44]. (A) Depicts the European cities' population. Instead of a cartogram, each city is represented as a circle with an area proportional to the city's population. *Source:* [Cleveland, 1985, p. 115] (B) Depicts the European empires' land mass after the 1801 Luneville Treaty and indicates that the British were paying more taxes than other nations. Circles represent the countries and their area represents the country's land mass. An Euler diagram is also provided at the bottom. The diagram depicts data in different forms: the population by a left vertical line to the circle; tax due by a right vertical line to the circle; relative tax burden on population by another line graph; maritime powers by colour; the empire's land mass proportion on the continents by pie charts. *Source:* [Tufte, 1983, p. 44]
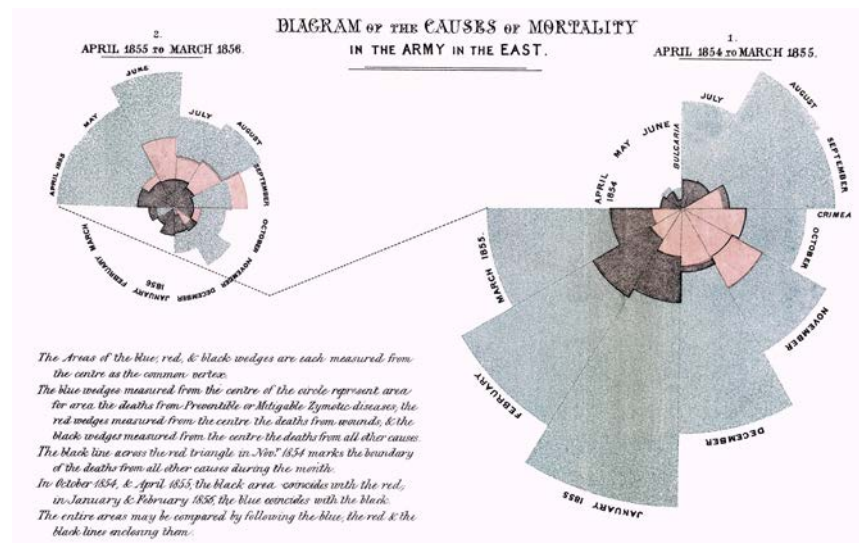
Figure 2.3: *First applications of area to represent quantities—Florence Nightingale.* Nightingale included this diagram (known as Nightingale's 'rose' or 'coxcomb' diagram) in her report [Nightingale, 1858] to the British government to demonstrate that during the Crimean War most of the soldiers died due to preventable diseases rather than wounds. The area of the radial segments in the diagram depict the number of deaths from preventable diseases (in blue), wounds (in red) and other causes (in black) per month for 24 months. Her presentation was so effective in communicating the issue that later the government implemented various reforms to improve the sanitary conditions in military hospitals. *Source:* [Nightingale, 1858]

Despite the use of data visualizations for all of these years, research in understanding how information should be effectively encoded in a visualization to facilitate communication and reasoning began more recently with Bertin's 'Sémiologie Graphique' [Bertin, 1983], first published in 1969. For the first time, the different ways how graphical representations encode data were systematically identified and various basic terms were defined. Bertin noted that every visualization is made up of basic units, known as *marks*, including points, lines and areas. These marks have properties, including position, size, shape, value, orientation, colour, texture and motion, which are used to represent the data and which are referred to as *retinal* or *visual variables*. Bertin also noted that the characteristics of these variables could be associative, selective, ordinal or quantitative in that marks are respectively perceived as similar, different, ordered or proportional, when changes in a specific variable are carried out (e.g., selection, order and quantity is perceived as the size of a mark be it point, line or area changes).

Following up on Bertin's work, Cleveland and McGill [1984; 1985; 1987] carried out experiments to identify which visual variables are most effective for specific data types and the accuracy at which humans can decode information encoded using these visual variables (referred to as elementary graphical-perception tasks). With their experimental results and theory of visual perception, they concluded that humans are best and most accurate at position judgement along a common scale, followed by position along non-aligned but identical scales, length, angle, area, volume and colour (hue, saturation, density) [Cleveland, 1994]. Thus, it is more difficult for humans to, for instance, identify differences in or quantify volume than length. These findings are important as visualization is not a process carried out by computers [Spence, 1996], but rather a 'cognitive activity' [MacEachren, 2004; Ware, 2012] carried out by humans, so that an effective visualization is one which acknowledges human capabilities and limitations [Spence, 2001].

Both Bertin's theoretical work and Cleveland and McGill's experimental findings are foundational to information visualization and will continue to provide guidance on the design of effective visualizations. To facilitate the design process, Mackinlay [1986a; 1986b; 1988] devised a system that used Bertin's semiology to automatically design effective visualizations for any data. In this way, Mackinlay managed to automate Bertin's theoretical work.

Other fundamental principles and good practices for the design of visualizations were provided by Tufte [1990; 1997b; 1997a; 2001; 2006]. Tufte emphasized the need for graphical excellence, in that concepts are "communicated with clarity, precision, and efficiency", the viewer is given "the greatest number of ideas in the shortest time with the least ink in the smallest space", and graphics tell "the truth about the data" [Tufte, 1983, p. 51]. He also illustrated the importance of graphical integrity, in that the depicted quantities on the graphic are directly proportional to the ones represented, text is added to the graphic to clarify the depicted data, and "data variation, not design variation" is shown [Tufte, 1983, p. 77]. Other practical design guidelines based on theories of human perception, cognition and attention were provided by Ware to direct designers in creating effective and useful visualizations that exploit the capabilities and manage the limitations of human psychology [Ware, 2008, 2012].

With such guidelines and advances in technology, various novel, sophisticated and specialized visual representations and visual interactions have been developed to analyse the vast amount of data that is routinely generated. In most application areas, relationships between data sets are studied for the understanding of processes, systems and communities, for informed decision-making and for knowledge discovery. Thus, over the years, different visual representations of data set relations have been proposed. We discuss such representations in the next section.

2.3 Visualizing Data Set Relations

"We are drowning in information, while starving for wisdom" [Wilson, 1998, p. 294]. Relationships between data sets are analysed to, for instance, understand biological processes and find cures to illnesses or to identify connections among communities and predict and disrupt crimes.

Often such data is structured as a graph and thus, over the years, various visual representations of graphs have been proposed. The two most popular are node-link diagrams and adjacency matrices. Figure 2.4A and Figure 2.4B are respectively the weighted node-link diagram and the weighted adjacency matrix in Schich's [2010, pp. 232-33] book chapter depicting the data model of the CENSUS database on Renaissance works of art and architecture [CENSUS, 2007]. So both visualizations depict the same data. Figure 2.4A is weighted as the size of the nodes and the thickness of the links indicate the number of nodes and links of that specific type. In Figure 2.4B, a line and a column for each node type is depicted and the number of links between every two types of nodes is noted in the cells of the matrix, so that the matrix is said to be weighted. Node-link diagrams tend to become cluttered very easily, as more nodes and links are added to the diagram. So, in such cases, adjacency matrices are often preferred. The effectiveness of such visualizations has been studied extensively (e.g., [Ghoniem et al., 2004; Keller et al., 2006]) and software systems that draw these visualizations and that aid users explore the visualized data have been devised (e.g., [Abello and van Ham, 2004; Henry et al., 2007; Henry and Fekete, 2006]). These have been used for various applications, such as social network analysis [Henry et al., 2007] and cancer data analysis and trend identification [Abello and van Ham, 2004]. Drawing techniques for depicting graphs as node-link diagrams are discussed in Section 2.4.

A

The CENSUS data model as a weighted node-link diagram
Maximilian Schich: Revealing Matrices. in: Beautiful Visualization. Looking at Data through the Eyes of Experts.
Ed. Julie Steele and Noah Ilinsky. Sabastopol, CA: O'Reilly. April 2010 fig. 3
URL: http://revealingmatrices.schich.info/fig/3.htm

B

	Document	Monument	Replica	Preservation Event	Provenance Event	Person	Location	Date	Style	Bibliography	Citation	Image	Inscription	Record History	Main Entry
total number of nodes:	31197	12688	566	2261	5878	1715	2317	3584	28	2421	36606	25865	37324	77746	17
Document	28869					6358	2706	5823		574	31391	24286	35996	38394	3087
Monument	43143	6579	578	2241	5885	6834	8226	5812	5033	27	9934	4653		15819	6451
Replica							113						88		566
Preservation Event	1008					1123	1527								2261
Provenance Event	9132						5867	5864							5878
Person						1666	3360							2409	1715
Location							2316							2221	1
Date														3688	3584
Style														35	28
Bibliography											7			1960	2414
Citation	see Figure 5										36592	4		3	2
Image														12663	25865
Inscription															37324
Record History															77746
Main Entry															24

Number of links.

The CENSUS data model as a weighted adjacency matrix
Maximilian Schich: Revealing Matrices. in: Beautiful Visualization. Looking at Data through the Eyes of Experts.
Ed. Julie Steele and Noah Ilinsky. Sabastopol, CA: O'Reilly. April 2010 fig. 4
URL: http://revealingmatrices.schich.info/fig/4.htm

Figure 2.4: *Visualizing the CENSUS data model using (A) a weighted node-link diagram and (B) a weighted adjacency matrix. Source:* [Schich, 2010, pp. 232-33]—Figure 14-3, Figure 14-4

It is also common for data to be structured as a tree (i.e., an acyclic connected graph) as is often the case with hierarchies (e.g., a file hierarchy). Various visual representations for these data structures have been proposed ([Schulz, 2011]—a visual survey). A few are radial visualizations ([Draper et al., 2009]—a survey), such as Calling Context Ring Charts (e.g., Figure 2.5A) [Moret et al., 2010], which

visualize Calling Context Trees containing dynamic metrics on the performance of object-oriented software. Both the caller and the callee methods are depicted as ring segments, with the segment of the callee placed around that of the caller. The area of the segments can also be proportional to a specific metric, such as the bytecode metric as in Figure 2.5A. Another visual representation is RINGS (e.g., Figure 2.5B) [Teoh and Kwan-Liu, 2002], which depicts all the nodes of the tree as circles and places the circles of the children nodes within the circle of the parent node. A colour map, as in Figure 2.5B, is then used to depict the number of children in the nodes.

Amongst the most popular tree visualizations are treemaps, which depict the nodes as rectangles with an area that is proportional to some attribute (e.g., the number of children nodes) and subdivide parent nodes to depict their children [Shneiderman, 1992]. Different variants of treemaps have been proposed. Squarified treemaps [Bruls et al., 2000], for instance, use rectangles that approximate as much as possible squares to avoid elongated rectangles. Cushion treemaps [van Wijk and van de Wetering, 1999] use a simple shading model to add ridges each time a rectangle is subdivided, so that the shading aids in depicting the hierarchical structure of the tree at different levels of details based on the selected height of the ridges. Cushion treemaps as well as squarified cushion treemaps, as Figure 2.5C, are used by the disk browser, SequoiaView [Bruls et al., 2002], to facilitate file management, file navigation and inspection of disk usage space.

All of these tree visualizations allow users to interactively navigate and explore the depicted data. A limitation is that intersections at the same level cannot be visualized and so, in a file system, for instance, a file cannot be in two directories unless one is a subdirectory of the other.
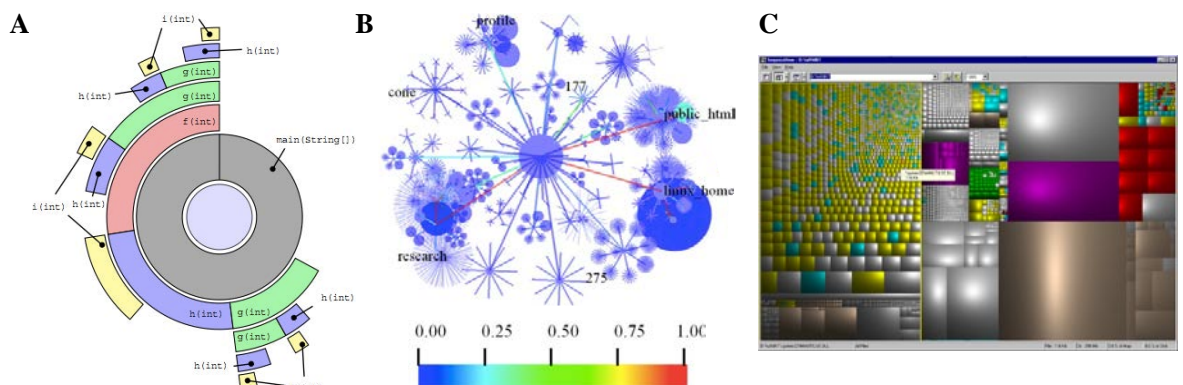
Figure 2.5: *Visualizing trees: (A) a Calling Context Tree using Calling Context Ring Charts, (B) a UNIX directory using RINGS, (C) a file hierarchical system using a squarified cushion treemap in SequoiaView. Sources:* (A) [Moret et al., 2010]—Figure 5; (B) [Teoh and Kwan-Liu, 2002]—Figure 7 (*left*); (C) [Bruls et al., 2002].

Circular visualizations, like Figure 2.6A, have also been proposed and software systems that automatically generate them have been devised (e.g., Circos [Krzywinski et al., 2009] which generated Figure 2.6A; CGView [Stothard and Wishart, 2005]; Radial Sets [Alsallakh et al., 2013]). Such layouts are thought to facilitate comparison of data sets, exploration and reasoning about relationships between the data elements. A more recently proposed visualization is LineSets [Alper et al., 2011] whereby each set is depicted as a coloured, curved line that passes through the elements of the set, as in Figure 2.6B.
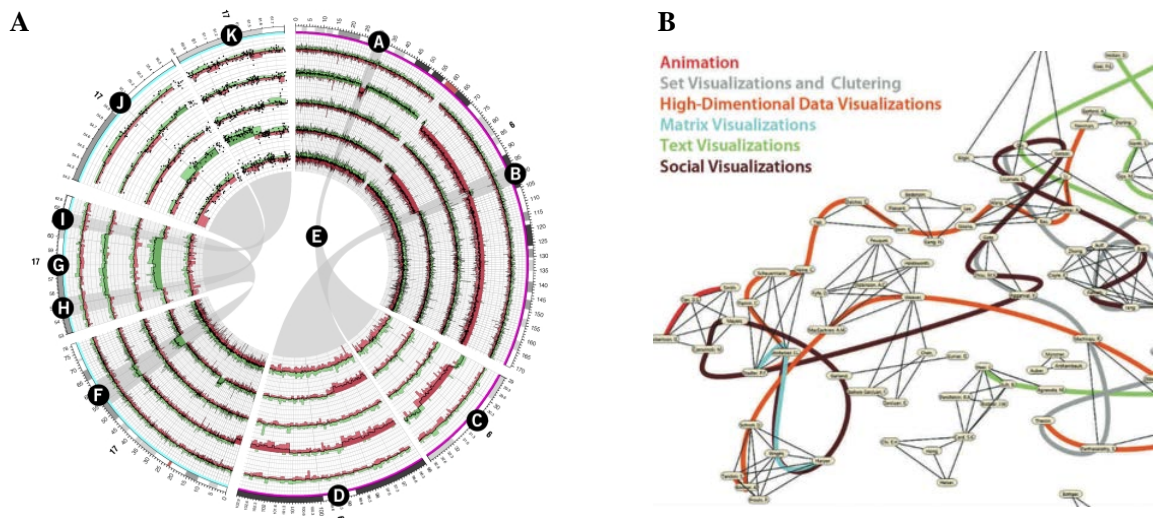
Figure 2.6: *Visualizing (A) genomic data using Circos's circular visualization, and (B) social network groups using LineSets. Sources:* (A) [Krzywinski et al., 2009] Figure 6; (B) [Alper et al., 2011] Figure 1.

In contrast to these visual representations, Venn and Euler diagrams, like those in Figure 2.7, depict the sets as closed curves, thus facilitating reasoning about the relations of the sets, as the sets and their overlaps are easily visible. This is so, as closed curves are preattentively processed [Treisman and Souther, 1985] and form common regions that have a strong perceptual grouping effect [Palmer, 1992]. Also, Euler diagrams can depict intersection, containment and exclusion amongst data sets all in one diagram (e.g., Figure 2.7A). The cardinalities of the sets and their intersections can also be depicted by the size of the curves and their overlaps (e.g., Figure 2.7B which in this case is a Venn diagram as it shows all the possible intersections between the curves). Venn and Euler diagrams are discussed in detail in the next chapter. Euler-like diagrams are discussed in Section 3.6.6.
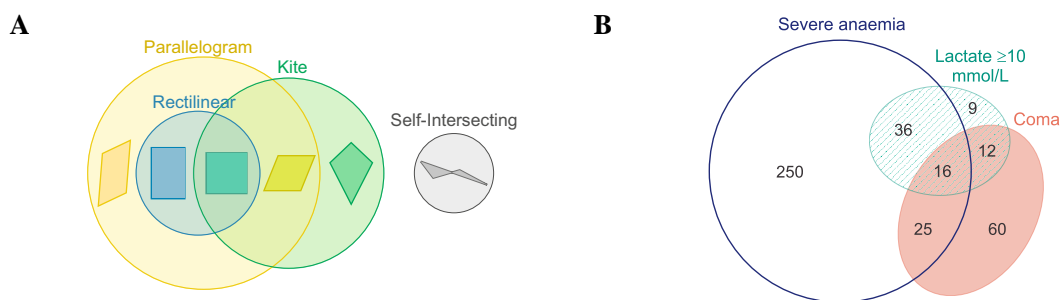
Figure 2.7: *Examples of an Euler diagram and a Venn diagram.* (A) An Euler diagram (drawn manually) classifying geometric shapes. (B) A Venn diagram whose curves and overlaps have area that is proportional to the data indicated by the numeric labels. The curves depict prognostic indicators of severe malaria (data source: [Dzeing-Ella et al., 2005]–Figure 3). The diagram was generated using eulerAPE (Chapter 5).

We will now discuss algorithms for the generation of node-link diagrams depicting graphs.

2.4 Graph Drawing

Various automatic techniques have been developed to draw graphs as node-link diagrams [Di Battista et al., 1994, 1999b; Díaz et al., 2002; Gibson et al., 2012; Sugiyama, 2002]. However, most techniques, such as those using simulated annealing [Coleman and Parker, 1996], genetic algorithms [Hobbs and Rodgers, 1998] or evolutionary algorithms [Utech et al., 1998], can be complex and computationally expensive. So lately, force-directed algorithms have been used [Gibson et al., 2012] to produce comprehensible graph layouts in relatively fast time [Kobourov, 2012]. Using a physical analogy, these methods are often simple, easy to reason about and to implement [Di Battista et al., 1999b].

2.4.1 Force-Directed Drawing Methods

The physical analogy used by force-directed algorithms is that of a system of physical structures (the vertices of the graph) that exert a force over others in the system, such that these structures move along the direction of the force [Di Battista et al., 1999a]. The system is brought to a halt when the algorithm positions the structures appropriately so that the forces are equilibrated.

A force-directed method is made up of: (i) *a force model* that is determined by the vertices, the edges and the forces exerted by each one; (ii) *an algorithm* that equilibrates the forces in the system. For such a method to generate an improved graph layout, the forces in the model should be defined appropriately so the required aesthetic criteria and constraints are met. Similarly, the algorithm should ensure that when the forces are equilibrated, the graph layout has been improved and satisfies the required criteria.

Various force-directed methods have been proposed to lay out graphs [Di Battista et al., 1999a; Gibson et al., 2012; Kobourov, 2012]). One of the very first is Tutte's [1960; 1963] barycentre method, that positions freely moving vertices at their neighbour's barycentre, that is the point between the vertex and the neighbour where the gravitational forces exerted by these two vertices are balanced.

However, one of the simplest methods that forms the basis of various others is the *spring embedder*, first proposed by Eades [1984], whereby the physical system is made up of electrically charged particles (the vertices of the graph) and springs (the edges connecting the vertices in the graph), as in the example in Figure 2.8A. So, while the particles repel each other, the springs attract the particles that are connected to their endpoints (Figure 2.8B). The forces in the system will move the particles around in space until their position equilibrates the forces (Figure 2.8C), at which point the improved graph layout should be obtained (Figure 2.8D). In this method, the force model is made up of a repulsive force between each pair of vertices and an attractive force between each pair of vertices that are connected with an edge. Typically, the repulsive forces are inversely proportional to the squared distance between the two vertices (consistent with the inverse-square law), while the attractive forces are directly proportional to the difference between the natural length of the springs in the physical system and the distance between the two vertices (consistent with Hooke's law) [Di Battista et al., 1999a]. So, while repulsive forces ensure that the vertices are not close to one another, attractive forces aid in obtaining uniform edge lengths [Di Battista et al., 1999a]. Often the vertices are evenly distributed and the graphs are symmetric [Eades and Lin, 2000], even though the force model does not directly take into account these features.
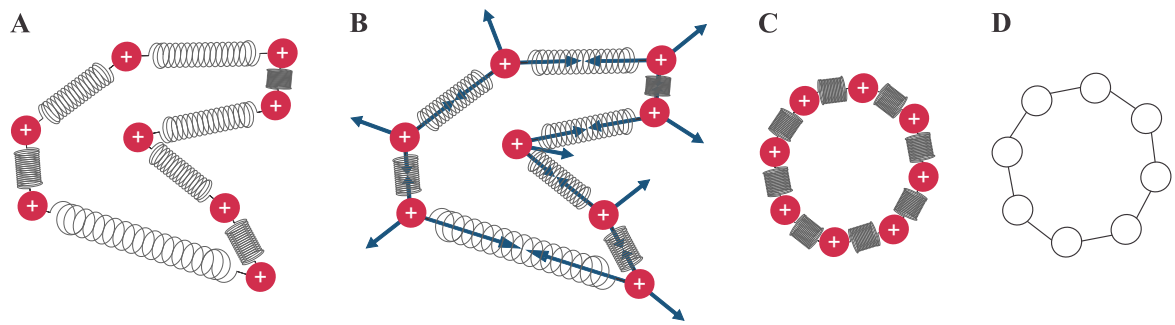
Figure 2.8: *The physical system and forces in a simple force-directed spring embedder.* (A) The vertices in the graph act like electrically charges particles, while the edges act like springs. (B) As the particles repel, the springs attract the particles attached to its endpoints. The vertices are moved around until (C) the position of the vertices equilibrates the forces in the system, at which point (D) an improved layout is obtained.

Various types of algorithms could be used to equilibrate the forces. A simple approach is to go through the system in discrete time steps, so that, starting with a graph layout whose vertices are randomly positioned, the forces exerted on the vertices are calculated at every step and the vertices are moved accordingly based on the magnitude and the direction of the forces [Eades, 1984]. The algorithm stops either after a fixed number of iterations or once the magnitude of the forces exerted on the vertices is zero. The transition from the initial to the final layout could be animated to facilitate understanding of how the forces in the system are acting upon each other [Di Battista et al., 1999a] and to possibly aid in preserving the mental map of the layout [Eades et al., 1991].

Alternative algorithms aim to avoid local minima and improve layout aesthetics by using numerical methods (e.g., [Kamada and Kawai, 1989]) or optimization heuristics (e.g., [Branke et al., 1997; Davidson and Harel, 1996]) to minimize the total energy of the system defined as a function in terms of the aesthetic criteria that should be satisfied by the final layout. In fact, different variants of Eades's spring embedder have been proposed (e.g., [Davidson and Harel, 1996; Fruchterman and Reingold, 1991; Hu, 2005; Kamada and Kawai, 1989; Sugiyama and Misue, 1995a, b]). For instance, magnetic fields have been added to the springs and the system to handle more aesthetic criteria (e.g., edge orientation) and to lay out other types of graphs (e.g., trees and directed graphs) besides undirected ones [Sugiyama and Misue, 1995a, b]. Others used Barnes and Hut [1986] algorithm to accurately and efficiently calculate the forces at every iteration to ensure that the final layouts satisfy the required criteria (e.g., [Hu, 2005; Tunkelang, 1999]).

Eades [1984] notes that his spring embedder is appropriate for non-dense graphs with fewer than 50 vertices. In fact, due to local minima, poor layouts are generated when graphs have hundreds of vertices [Kobourov, 2012]. So a few variants (e.g., [Gajer et al., 2001; Hadany and Harel, 2001; Harel and Koren, 2001, 2001; Hu, 2005; Walshaw, 2001]) adopted a multilevel approach, whereby vertices are iteratively clustered to form coarser and coarser graphs until a graph with a small number of vertices is obtained, at which point, layout optimization commerces with the coarsest graph and continues with others, until the original graph is reached. This approach has been successful in overcoming local minima and in appropriately and efficiently laying out large graphs. For instance, using this approach, Hu's [2005] (also [2011]) method lays out graphs with over 10,000 vertices in less than a minute.

Various force-directed algorithms are discussed and compared in Kobourov's [2012] review and in Gibson et al.'s [2012] survey. Such methods have also been used to bundle edges in graphs [Holten and

van Wijk, 2009], thus reducing visual clutter, and to lay out diagrams, such as 3D UML class diagrams [Dwyer, 2001]. This indicates their applicability to other areas besides graph drawing. Other graph layout methods and aesthetics of node-link diagrams for graphs are briefly discussed in Appendix C.1.

Visualizations could facilitate data analysis and reasoning, particularly in areas that are notorious for fallacious reasoning. Decision-making with statistical, uncertain data that involves Bayesian inferencing is an example. We discuss this further in the next section.

2.5 Visualizations for Bayesian Reasoning

Both laymen and professionals have difficulty making inferences and decisions based on statistical and probabilistic data and often they are unaware of their wrong judgements [Gigerenzer et al., 2007; Hoffrage et al., 2000]. In cases of uncertainty, judgement could also be affected by gut feelings [Gigerenzer, 2007], past experiences, emotions [Slovic et al., 2005], and whether making decisions for oneself or others [Zikmund-Fisher et al., 2006]. This can have severe consequences in many domains.

2.5.1 A Few Bayesian Reasoning Scenarios

Physicians need to diagnose diseases based on the outcome of unreliable medical tests. Patients need to decide whether they should undertake heavy medical treatment. Wrong judgements are common and often result in overdiagnosis [Welch and Black, 2010]—example, up to two thirds of the breast cancers that are detected by mammography screening can be overdiagnosed [Zahl and Maehlen, 2004]. In other cases, patients with a positive HIV test result attempted or committed suicide before further tests turned out negative [Chicago Tribune, 1993; Stine, 1998]. In this domain, a crucial piece of information for effective decision-making is the probability that a patient has a disease given that a test is positive.

In legal trials, juries have to convict or acquit defendents based on unreliable evidence and here too, wrong judgements abound [Koehler, 1997]. A Harvard law professor, Alan Dershowitz, who advised defence lawyers in a murder trial of a husband who battered his wife and whose wife was murdered claimed on US television that only 0.1% of husbands who batter their wives end up murdering them [Good, 1995]. However, Dershowitz's reasoning was fallacious, as the probability that the husband was the murderer is conditional on both the events that he battered his wife and that she was killed (giving a probability of 50%) rather than just the event that he battered her (giving 0.1%) [Good, 1995].

Intrusion detection systems are not 100% effective at detecting intrusions as decisions are based on uncertain information [Axelsson, 2000]. False alarms can be costly and intrusions can be disregarded. So, the chance that a real intrusion has been identified when the alarm goes off has to be determined. Other scenarios in everyday are provided by Hastie and Dawes [2001].

2.5.2 The Base Rate Fallacy

The scenarios in Section 2.5.1 involve Bayesian inference, which is known to be counterintuitive and subject to fallacious reasoning. As an illustration, consider the following classic Bayesian problem:

> The probability that a woman at age 40 has breast cancer is 1%. According to the literature, the probability that the disease is detected by a mammography is 80%. The probability that the test misdetects the disease although the patient does not have it is 9.6%.
>
> If a woman at age 40 is tested as positive, what is the probability that she indeed has breast cancer?
>
> <div align="right">[Eddy, 1982]</div>

Out of 100 doctors, 95 estimated this probability to be between 70% and 80% [Eddy, 1982], but the correct probability is only 7.8% (Appendix C.2.1 explains how this is computed using Bayes' theorem). The probability is low because the prevalence of the disease in the population, that is the base rate (here 1%), is low. When making Bayesian inference, this information is often ignored [Gigerenzer and Hoffrage, 1995], leading to the base rate fallacy [Barbey and Sloman, 2007a, b]. Studies suggest that the presentation of the problem is decisive for the use or neglect of the base rate [Gigerenzer et al., 1988]. Presentations with natural frequencies (Section 2.5.3) and visualizations (Section 2.5.4) have been used.

2.5.3 Using Natural Frequencies

Natural frequencies could be used, so that the Bayesian problem in Section 2.5.2 is presented as:

> 10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.
>
> Here is a new representative sample of women at age forty who got a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer?
>
> <div align="right">[Gigerenzer and Hoffrage, 1995]</div>

Using natural frequencies instead of probabilities reduces the fallacy [Hoffrage et al., 2002], with a typical improvement that goes from 16% to 46% of correct answers (over 15 problems and 60 university students, most from psychology) [Gigerenzer and Hoffrage, 1995]. Similar improvements were observed in other studies (e.g., [Brase, 2008; Cosmides and Tooby, 1996]). Appendix C.2.2 discusses why natural frequencies reduce the fallacy. Yet a success of 46% is still low. So other representations that aid understanding of how the base rate (here 1% or "10 of every 1,000"), the hit rate (here 80% or "8 of every 10") and the false alarm rate (here 9.6% or "95 out of every 990") of the problem are related, should be studied for improved success rates. Visual representations could help.

2.5.4 Using Visualizations

To solve Bayesian problems without using Bayes' [1763] theorem, heuristics (e.g., [Kahneman et al., 1982]), cognitive algorithms (e.g., [Gigerenzer and Hoffrage, 1995]) and other methods that do not use visualizations (Appendix C.2.3) have been used. Solutions that use visualizations were also proposed as

diagrams are known to facilitate reasoning [Bauer and Johnson-Laird, 1993]. In fact, a study confirms that when Bayes' theorem is introduced to students through visualizations, students learn faster and report higher temporal stability than without a visualization [Sedlmeier and Gigerenzer, 2001]. Prior training is not always possible and so, a few studies were conducted to assess the immediate benefits of visualizations (e.g., [Brase, 2009; Cole and Davidson, 1989; Cole, 1989; Sloman et al., 2003]). However, it is still unclear which is the most effective representation for Bayesian reasoning.

The visual representations that have been considered for Bayesian reasoning include: contingency tables [Cole and Davidson, 1989; Cole, 1989], detection bars [Cole and Davidson, 1989; Cole, 1989], signal detection curves [Cole and Davidson, 1989; Cole, 1989], Bayesian Boxes [Burns, 2003, 2004a, b], probability space diagrams [Cheng and Pitt, 2003], frequency trees [Sedlmeier and Gigerenzer, 2001], and probability trees [Sedlmeier and Gigerenzer, 2001]. Examples of such visualizations are available in Appendix C.2.3 Figure C.2. Most of these representations attempt to visualize the structure of the problem and how the different quantities (i.e., the base rate, the hit rate and the false alarm rate) relate, but most are difficult to understand, thus require training prior to use.

Two popular visualizations are Euler diagrams and frequency grids. Studies suggest that Euler diagrams (e.g., Figure 2.9A) can clarify the nested-set relations and categorical information of Bayesian problems (how the quantities relate) [Sloman et al., 2003], while frequency grids (e.g., Figure 2.9B) can facilitate logical reasoning [Brase, 2008; Cosmides and Tooby, 1996].
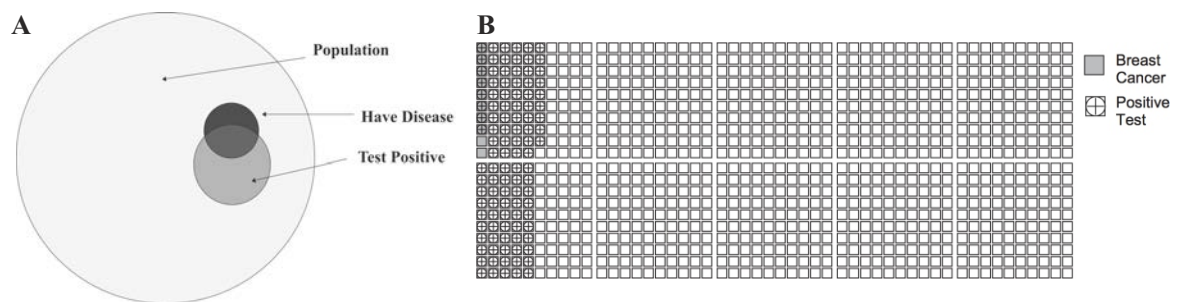
Figure 2.9: *An Euler diagram and a frequency grid for Bayesian reasoning to avoid the base rate fallacy.* (A) An Euler diagram (for a general disease diagnosis problem [Brase, 2009]). *Source:* [Brase, 2009]–Figure 1, top diagram (B) A frequency grid (for the mammography problem [Eddy, 1982]). *Source:* [Sedlmeier and Gigerenzer, 2001]–Figure 5, top section

Part of this dissertation evaluates Euler diagrams, frequency grids and combinations of both for Bayesian reasoning. Studies on Bayesian reasoning assessing frequency grids are discussed in the next section, while others assessing Euler diagrams or combinations of both are discussed in Section 3.9.

2.5.5 Using Frequency Grids

Cosmides and Tooby [1996] argued that representations with discrete, countable objects like frequency grids (e.g., Figure 2.9B, Figure 2.10) facilitate logical reasoning and understanding of how categories relate, as the human mind is adapted to process information in natural frequency format. However they found no improvement over text alone (success of 76% of 75 participants with frequency grids and 76% of 50 participants with text alone, using natural frequencies, over a major revision of Casscells et al.'s

[1978] disease diagnosis Bayesian problem; participants were paid Stanford University students with 19.6 years average age). They believed that visualizations were ignored, as they observed a notably improved success of 92% of 25 participants when guided to actively draw their own frequency grid.

Similarly, after training, Sedlmeier and Gigerenzer [2001] observed a success of 75% then 100% five weeks later among 14 participants who drew frequency grids, compared to 60% then 20% (as before training) among 20 participants applying Bayes' theorem (participants were paid university students and 10 Bayesian problems were tested). However later, Brase [2009] reported no improvement between passive and actively drawn grids (49% success of 49 participants for active, 48.4% of 95 for passive, over one problem similar to that of Cosmides and Tooby [1996]; participants were university Psychology students with 18.7 years average age, who participated to partly fulfil a class requirement).

Cole and Davidson [1989] (see also [Cole, 1989]) showed that subjects become remarkably accurate and fast when trained using a frequency grid (or any of the other two visualizations they tested) rather than a contingency table and compared to the use of text alone with no training. However, differences in errors between frequency grids and other visualizations were not significant (18 Psychology students from Carnegie Mellon University per visualization participated to fulfil a course requirement; eight problems with diverse 'high'/'low' value combinations for the base, hit and false alarm rates were tested).

Brase [2009] also compared frequency grids with regular (Figure 2.10A) and random (Figure 2.10B) layouts, but found no difference in success rate (47.6% of 42 university Psychology students for both with 18.8 years average age, over one problem similar to Cosmides and Tooby's [1996]).
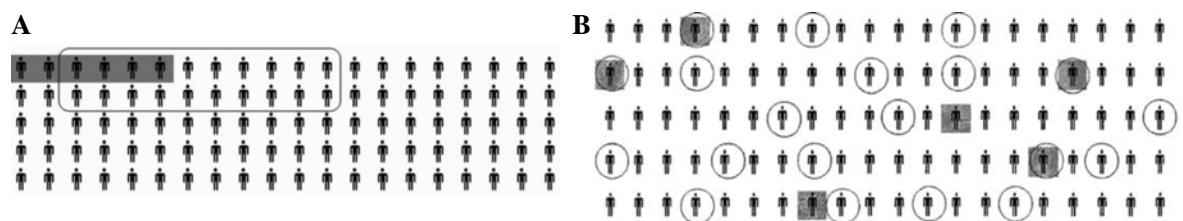
Figure 2.10: *Brase's [2009] frequency grids with (A) a regular layout and (B) a random layout. Source:* [Brase, 2009]—(A) Figure 1, bottom diagram; (B) Figure 2, top diagram

Kellen et al. [2007] argued that the effectiveness of the visualization is determined by the user's spatial abilities. Those with high spatial abilities can handle any visualization particularly those with spatial properties, while those with low spatial abilities have difficulties handling visualizations with spatial properties and are better at handling ones with discrete objects [Kozhevnikov et al., 2002, 2005]. This led to Kellen et al.'s hypothesis that visual representations with icons like frequency grids (Appendix C.2.4 Figure C.4) are more effective for Bayesian reasoning than ones with spatial properties like Figure 2.9A. However, the findings of their proposed study were not published.

Frequency grids with tree diagrams [Spiegelhalter et al., 2011] and frequency grids with interactive features [Tsai, 2012] have also been proposed to facilitate Bayesian reasoning. Such visualizations are not the focus of this research work and are thus discussed in Appendix C.2.5 and Appendix C.2.6 respectively. Further reading on frequency grids is provided in Appendix C.2.7 – Appendix C.2.9.

Proposed visualizations for Bayesian reasoning are provided with the text conveying the problem. Guiding the user to draw their own diagram [Cosmides and Tooby, 1996] or using interactive visualizations [Tsai, 2012] could help as the users experience the inferencing process [Sedlmeier, 1999].

However, these solutions are not practical as often information is provided on paper (e.g., medical pamphlets, newspapers) and timely decisions have to be made (e.g., military generals in war). So techniques that holistically combine text and visualization should be considered. We discuss a few techniques in the next section.

2.6 Visualization and Text

Visualizations can facilitate understanding of information that is often conveyed using text, but they have to be linked to the text appropriately for the reader to benefit from their availability [Tufte, 2001].

Consider the two representations for part of Euclid's Pythagorean theorem proof in Figure 2.11. Figure 2.11A is a classic representation that uses letters to refer to elements of the diagram in text. Figure 2.11B is Byrne's [1847] visual representation that uses colours and shows the actual elements (i.e., parts of the diagram) in the text, sometimes as nouns instead of words. The link between the diagram and text is easily noticeable in Figure 2.11B and facilitates understanding. In Figure 2.11C, Tufte [1990, p. 87] added letters to Figure 2.11B and argued that, even though redundant symbols are often confounding, in this case, the letters rapidly link the diagram to the proof as they facilitate eye movements between the two. Thus, it is likely that both representations would be used together, but if not, the readers can choose their preferred way of linking the diagram with the text [Tufte, 1990, p. 87].
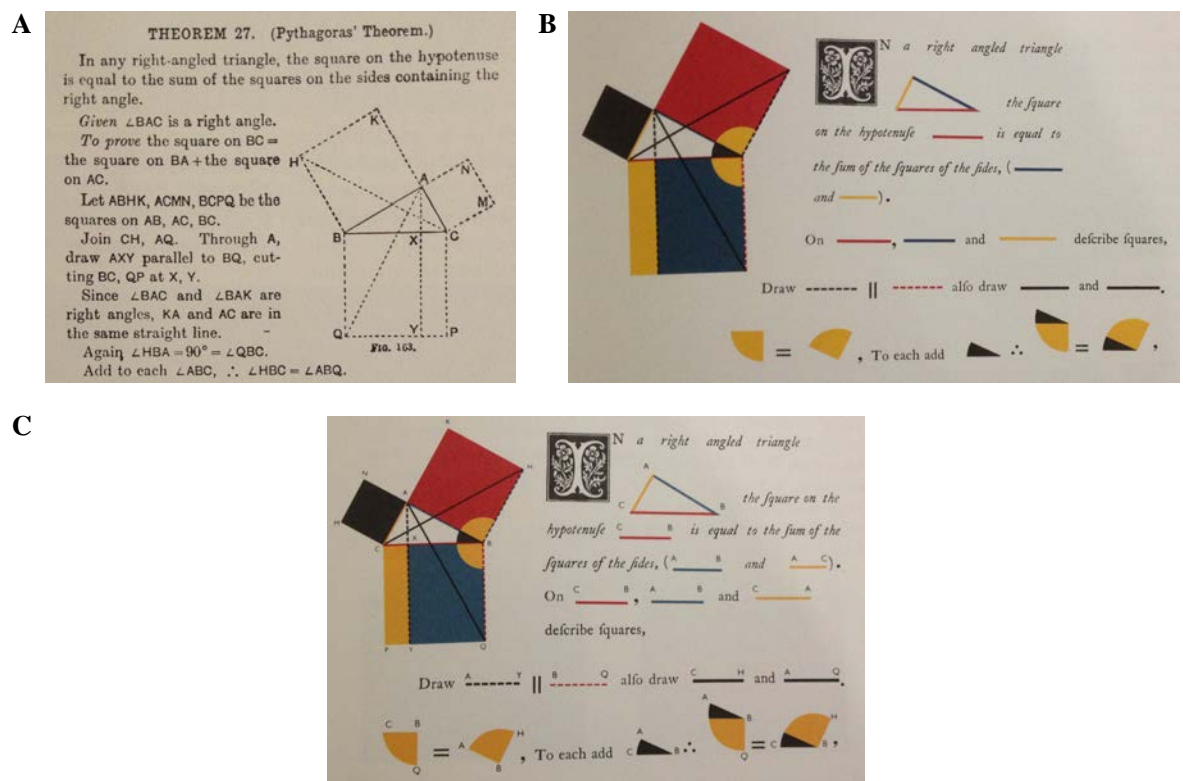
Figure 2.11: *Different representations for part of Euclid's Pythagorean theorem proof.* (A) A classic representation using labels. *Source*: [Tufte, 1990, p. 84]; original source: [Durell, 1936, p. 119] (B) Byrne's representation using colours and the actual elements of the diagram in the text in his visual edition of Euclid's *Elements* [Heath, 1956]. *Source:* Tufte's [1990, p. 85] redrawing of the original [Byrne, 1847, pp. 48-49] (C) Byrne's hybrid representation combining A and B. *Source:* [Tufte, 1990, p. 87]

A classic in information visualization that "exemplifies many of the fundamental principles of analytical design" [Tufte, 2006, p. 126] and illustrates how to "Completely integrate words, numbers, images, diagrams" [Tufte, 2006, p. 126] is Minard's [1869] flow map (Figure C.8) showing the loss of men in Napoleon's army as they moved from Poland to Russia and as they later retreated, during their invasion of Russia between 1812 and 1813. Besides labelling all the interesting and necessary data, Minard also includes some textual information to describe the encoding mechanism (e.g., the widths of the coloured regions depict the number of men), the measurement scales, the source of the data and any assumptions made. Though at first sight there seems to be a lot of words, there is no cognitive overload as the provided details are necessary to understand the data and the story (without the need to read about the invasion) and "the map's essential features remain in our memory long after we have ceased looking at it" [Spence, 2007, p. 2]. Minard's graphic was meant for "exploration" rather than "communication and illustration of a settled fact" [Tufte, 1983, p. 182] and so, he rightly used text to instruct the reader on "*how* to read the design … and not *what* to read in terms of content" [Tufte, 1983, p. 182].

Both visualization and text present information and so, "Data graphics are paragraphs about data and should be treated as such" [Tufte, 1983, p. 181] (Tufte's "data/text integration" principle). Further reading on techniques to combine visualization and text is provided in Appendix C.3.

2.7 Summary

Information visualization has developed extensively over the years. A number of visualizations have been proposed to facilitate data analysis and reasoning in diverse areas. Visualizations could facilitate Bayesian reasoning by depicting the structure of the problem and how different data sets relate. This could reduce fallacy reasoning that could otherwise have severe consequences in areas like medicine, law and military. The context of a visualization has to be defined for it to be meaningful. Pictures alone cannot define the context and text is often required. Even so, a visualization is helpful, only if the text and the visualization are integrated appropriately and the link between the two is clear.

In contrast to other visual representations, Euler diagrams represent various relationships between data sets in one diagram, including containment, exclusion and intersections. If all the possible overlaps between the data sets are represented, the Euler diagram is a Venn diagram. Venn and Euler diagrams are extensively used for various applications areas and are the main focus of this dissertation. The next chapter provides an overview of the current state-of-the-art of Venn and Euler diagrams and discusses areas that need to investigated further.

Chapter 3

Venn and Euler Diagrams

This chapter provides a literature review on work related to Venn and Euler diagrams. This will help the reader understand the current state-of-the-art and concepts discussed in this dissertation.

3.1 Introduction

Data sets in diverse areas are compared and their relationships are analysed. For instance, in genetics, relationships and differences between groups of genes and proteins are studied to understand biological processes, find cure for illnesses, and solve problems in agriculture. In national security, connections between communities in social networks are investigated to predict and disrupt criminal threats.

To facilitate analysis, Venn and Euler diagrams are often used to depict relationships between data sets. These diagrams are made up of labelled closed curves (of any shape) each depicting a data set (Figure 3.1). As the curves intersect, regions depicting data set overlaps are formed. An Euler diagram (Figure 3.1A) can represent any containment, exclusion and intersection of the data sets in one diagram [Euler, 1768a], but a Venn diagram (Figure 3.1B) depicts all the possible intersections of the data sets [Ruskey and Weston, 1997; Venn, 1880]. The properties of these diagrams are discussed in Section 3.4.
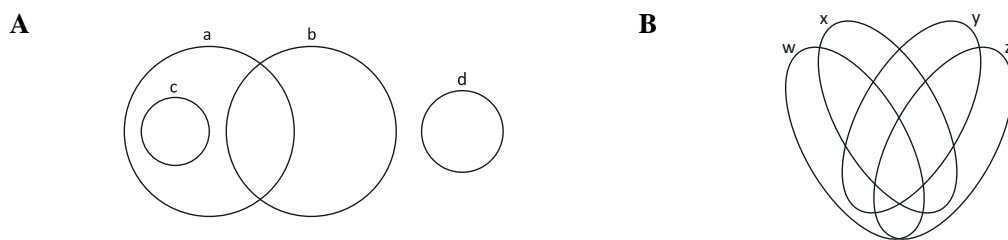
Figure 3.1: *An Euler diagram (left, A) and a Venn diagram (right, B), with four curves.* (A) Depicts containment, exclusion and intersection, as curve *c* is contained in *a* (so *c* is a subset of *a*), curves *a* and *b* intersect, and curve *d* is excluded. (B) Depicts all the possible intersections between curves *w*, *x*, *y* and *z*.

Other visualizations for sets and their relations are available (Section 2.3), but these alternatives do not use closed curves to represent data sets and so, overlaps between the sets are not easily visible. The

perceptual grouping principle of common regions [Palmer, 1992] indicates that elements within an enclosed region or curve tend to be grouped together. This grouping effect is stronger than and independent from any others like the Gestalt powerful proximity and similarity [Koffka, 1935]. Closed curves [Enns, 1986; Treisman and Souther, 1985] and crossing line segments [Julesz and Bergen, 1983] are also processed preattentively, so the curves and regions are distinguishable and pop out as discrete complete objects. Closure is also a Gestalt perceptual grouping principle [Koffka, 1935]. These factors motivated Ware's [2012] design guideline of "putting related information inside a closed contour".

Also, most of these alternative representations are not capable of visualizing containment, exclusion and intersection relations all in one diagram, as is possible with Euler diagrams. This allows Euler diagrams to be "well-matched to meaning" [Gurr and Tourlas, 2000] and the set relations they depict, thus facilitating reasoning [Stenning and Oberlander, 1995] and providing facts for free (known as 'free-rides' [Shimojima, 1996, 2004] or 'cheap rides' [Gurr, 2001]) without inferencing. For an example of a 'free-ride', consider Figure 3.2A. We know that *c* is a subset of *a*. So, if a curve *e* is placed in *c* as in Figure 3.2B, then *e* is a subset of *c*. However, since *c* is a subset of *a* then for free we know that *e* is also a subset of *a*. This partly explains why these diagrams are used for reasoning systems [Stapleton, 2005].
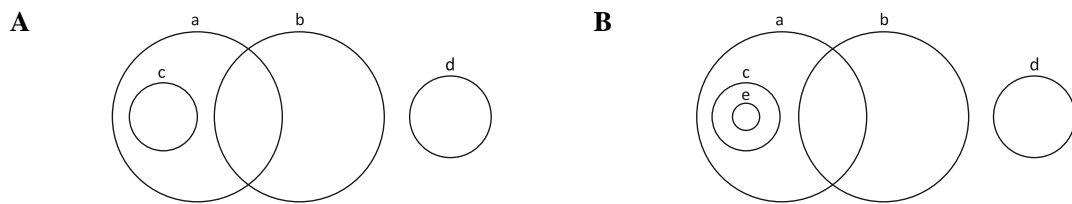
Figure 3.2: *An example of a 'free-ride'.* (A) Curve *c* is a subset of curve *a*. (B) Curve *e* is added to diagram A so *e* is a subset of *c*. Since *c* is a subset of *a*, then we know for free that *e* is also a subset of *a*.

This review outlines the state-of-the-art of Venn and Euler diagrams based on concepts discussed in this dissertation. Historical background (Section 3.2), application areas (Section 3.3), properties of these diagrams (Section 3.4) and related aesthetic and perceptual factors (Section 3.5) are discussed, followed by current automatic drawing techniques for Venn and Euler diagrams (Section 3.6) whose curves only depict the data sets. Sometimes the area of the regions is proportional to the depicted quantitative data. Such area-proportional Venn and Euler diagrams are discussed (Section 3.7), including their specific application areas (Section 3.7.1), properties (Section 3.7.2), relevant aesthetic and perceptual factors (Section 3.7.3), automatic drawing techniques (Section 3.7.4), theoretical findings (Section 3.7.5), and difficulties in drawing these diagrams (Section 3.7.6). Instead of area, glyphs that are proportional in number could be placed in the regions to depict associated quantitative data. Euler diagrams with glyphs and drawing techniques for these diagrams are discussed (Section 3.8), followed by an overview of how Euler diagrams are used for Bayesian reasoning (Section 3.9). We conclude with a discussion of how the contributions of this dissertation try to resolve current important issues in the area (Section 3.10).

3.2 Historical Background

Euler diagrams are named after the Swiss mathematician and physicist Leonhard Euler (1707-1783). Between 1760 and 1762, the King of Prussia hired Euler to teach his niece, the Princess of Anhalt-Dessau, science and philosophy [James, 2002]. His 234 letters with lessons for the King's niece, known as the 'Lettres à une Princesse d'Allemagne' [Euler, 1768b], became world renowned, widely translated

and "one of history's finest examples of popular science" [Dunham, 1999, p. xxv], which laymen could understand. To teach categorical propositions and syllogisms [Euler, 1768a], Euler used overlapping circles as in Figure 3.1A. In Figure 3.1A, it is easily noted that all *b* are *a*. So, if for instance *a* depicts vertebrates and *b* fish, then we know that all fish are vertebrates. These diagrams were later referred to as Euler diagrams or Euler circles [Baron, 1969]. In one of his letters, Euler claims:

> These circles, or rather these spaces … are extremely commodious for facilitating our reflections on this subject, and for unfolding all the boasted mysteries of logic, which that art finds it so difficult to explain; whereas, by means of these signs, the whole is rendered sensible to the eye.
>
> [Euler, 1795, pp. 453-54]

Baron [1969] notes that Euler was not the first to use such diagrams. German mathematician and philosopher Gottfied Wilhelm Leibniz (1646-1716) used Euler-like circles for logical reasoning [Couturat, 1901], but his work was not published [Couturat, 1903]. Similar diagrams were drawn by Ramon Lull (1235-1315) to reason about combinations of concepts [Lull, 1617]. Yet Euler was the first to make Euler diagrams accessible to all (also laymen) through his letters to the princess [Euler, 1768a].

Over a hundred years after Euler's letters, a logician, philosopher and Cambridge priest John Venn introduced Venn diagrams and a shading mechanism to indicate which of the set overlaps do not exist in the actual data [Venn, 1880]. So while an Euler diagram shows only the overlaps of interest, a Venn diagram shows all the possible set overlaps and then shades those that do not exist in the actual data. Thus the Euler diagram and the Venn diagram in Figure 3.3 are representing the same data set relations.
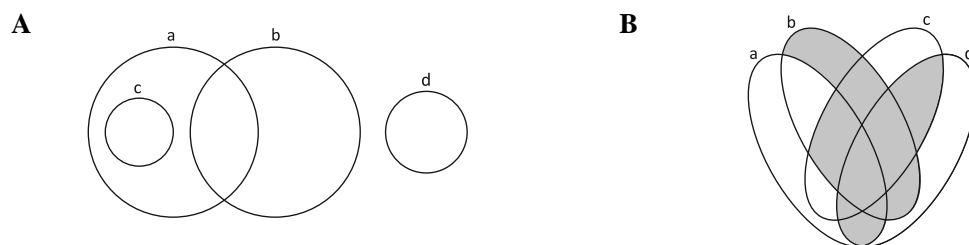
Figure 3.3: *An Euler diagram (left, A) and a Venn diagram (right, B) representing the same data set relations.* Regions in B that are not depicted in A are shaded, for B to represent same data set relations as A.

3.3 Application Areas

Venn and Euler diagrams are used in various applications areas for different types of data sets. Figure 3.4 illustrates a few general examples. More specific application areas include: biosciences and genetics (e.g., Figure 3.5, Figure 3.6A, [Kestler et al., 2008; Lim et al., 2005]); medicine and public health (e.g., [Carr, 1996; Dyer et al., 2007]); psychology (e.g., [Henson, 2006; Schubert, 1999]); ecology (e.g., [Coreau and Martin, 2007; Liao et al., 2011]); military (e.g., [Laabs and Baker, 1989]); large database querying (e.g., [Verroust and Viaud, 2004]); computer file systems (e.g., [Cordasco et al., 2012; De Chiara et al., 2003]); robotics (e.g., [Quick et al., 2006]); ontologies (e.g., [Howse et al., 2011]); Bayesian reasoning (Section 3.9); reasoning systems (e.g., spider diagrams—Figure 3.6B; constraint diagrams [Kent, 1997]; see Stapleton's [2005] survey); theorem provers [Urbas et al., 2012; Urbas and Jamnik, 2011, 2012]. Euler-like curves have also been used to group fixed points [Collins et al., 2009; Dinkla et al., 2012] and nodes in 3D structures [Balzer and Deussen, 2007].
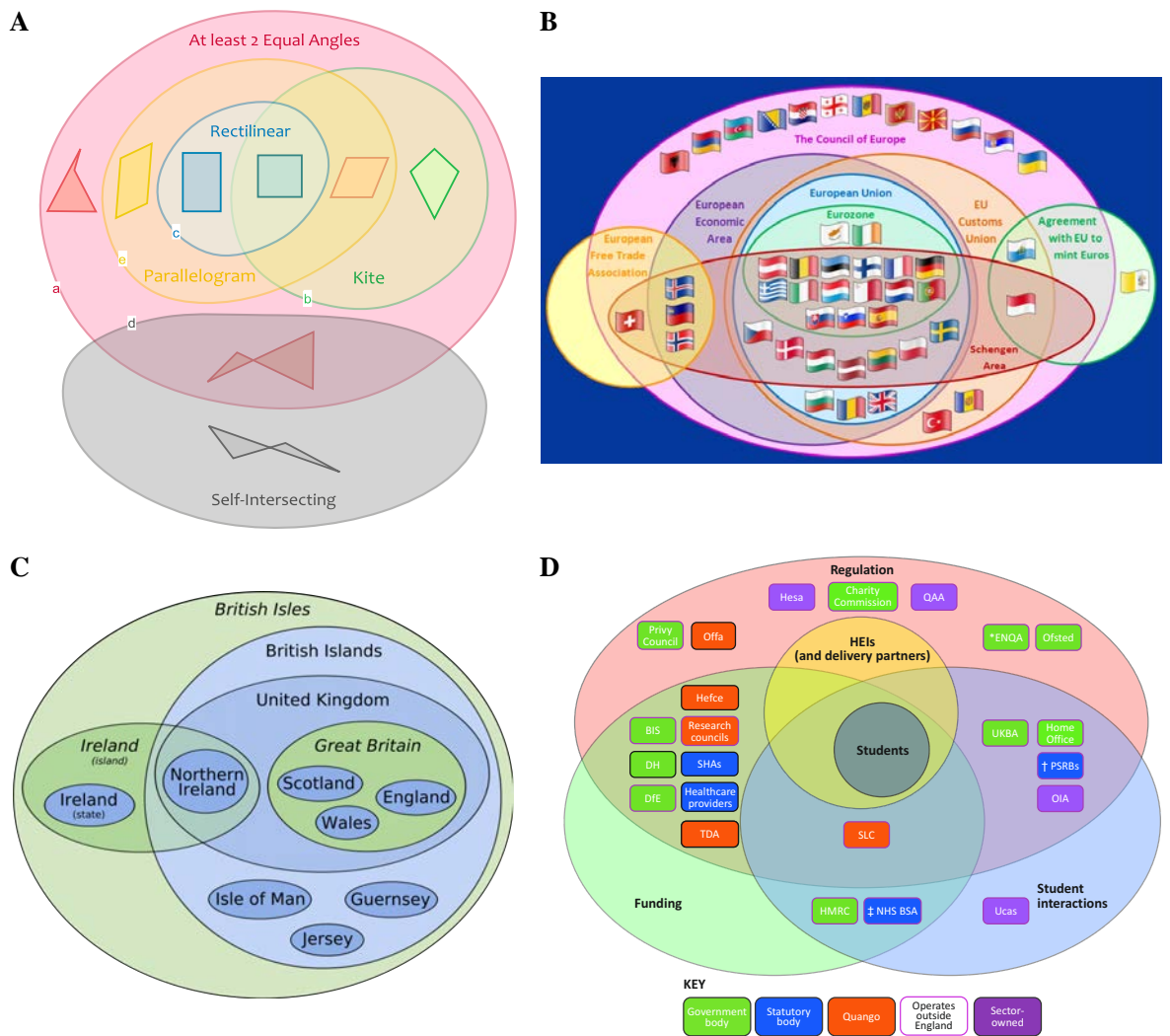
Figure 3.4: *Examples of Euler diagrams.* (A) Classification of shapes based on their properties. *Source:* [Micallef and Rodgers, 2009] poster (B) European bodies, their members and relations. *Source:* [Wikimedia Foundation, 2013] (C) The British Isles. Green for geographical features, blue for legal distinctions. *Source:* [Wikimedia Foundation, 2011] (D) UK's higher education system. *Source:* [Morgan, 2012]—redrawn
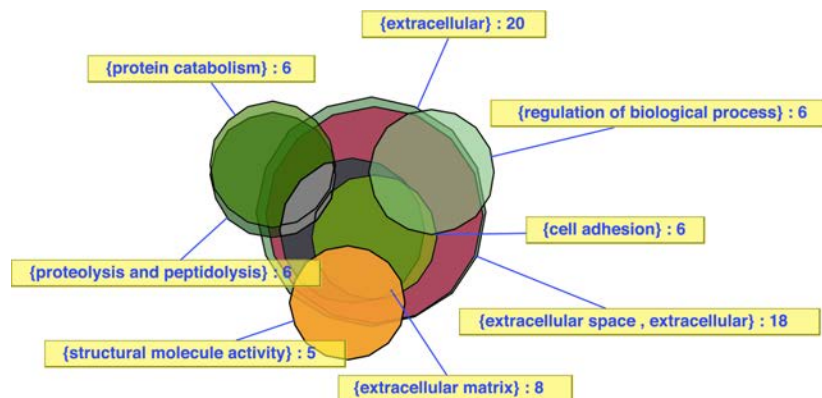
Figure 3.5: *An Euler diagram in genetics showing the results of a microarray experiment. Source:* [Kestler et al., 2008]–Figure 3

A

B
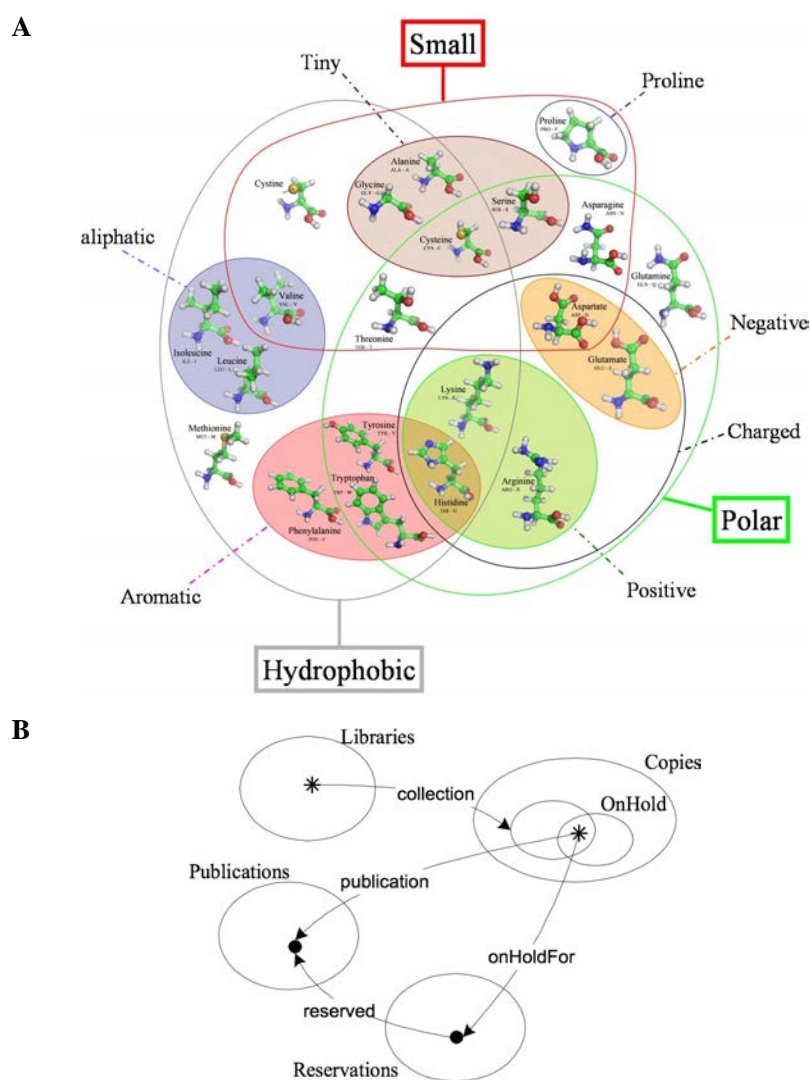
Figure 3.6: *Examples of Euler diagrams in biosciences and logic.* (A) Properties of amino acids. *Source:* [Podtelezhnikov, 2008] (B) A spider diagram for a library system. *Source:* [Howse et al., 2001]–Figure 1.1

3.4 Properties of a Venn Diagram and an Euler Diagram

Venn and Euler diagrams are made up of labelled closed curves (Section 3.1; e.g., Figure 3.1). Each closed curve splits up the plane into two components, one interior to the curve and another exterior to the curve. When a number of closed curves intersect, the interior of each curve is split up into smaller components, each of which is completely bound by segments of the curves. Excluding the components that are exterior to all of the curves, all the other components will be part of the interior of one or more curves. Every connected component in the diagram is a *minimal region* of the diagram. The set of minimal regions located in exactly the same set of curves makes up a *zone* of the diagram. The set of minimal regions that are exterior to all of the curves in the diagram is also a zone of the diagram.

A zone made up of just one minimal region is referred to as *connected*, while a zone made up of more than one minimal region is referred to as *disconnected*. Figure 3.7A has eight minimal regions and eight zones all of which are connected, while Figure 3.7B has ten minimal regions and eight zones of which two are disconnected. Formal definitions of all properties are provided by Stapleton et al. [2007].
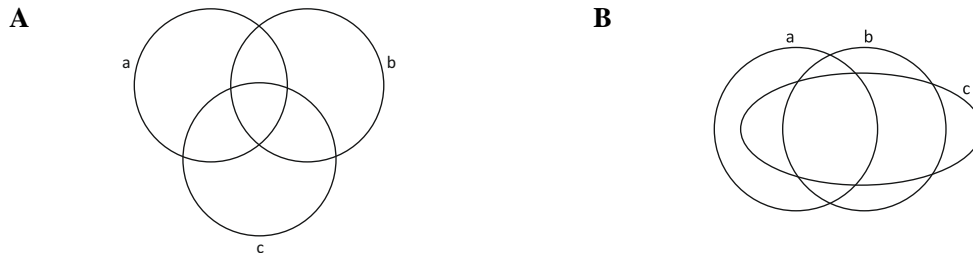
Figure 3.7: *Venn diagrams with connected (left, A) and disconnected (right, B) zones.* (A) All the zones are connected, as all are made up of one minimal region. (B) Two zones are disconnected, as there are two minimal regions inside curves *a* and *b* but not *c*, and two minimal regions inside curve *b* but not *a* and not *c*.

An Venn diagram with *n* curves is referred to as an *n-Venn diagram* or just an *n*-Venn. Similarly, an Euler diagram with *n* curves is referred to as an *n-Euler diagram* or just an *n*-Euler. So Figure 3.7A and Figure 3.7B are 3-Venn diagrams, Figure 3.1A is a 4-Euler diagram, Figure 3.1B is a 4-Venn diagram.

3.4.1 The Curves

The curves in the diagram can have any shape, unless some constraints are imposed. Examples of such constraints include: curves should have a particular geometric shape such as a circle (e.g., [Stapleton et al., 2011c, 2012]), an ellipse (e.g., [Stapleton and Rodgers, 2011]; the work in Chapter 5), a triangle (e.g., [Carroll, 2000; Swinton, 2009]), a rectangle (e.g., [Marshall, 2005]), a rectilinear shape (e.g., [Chow and Ruskey, 2004]), a regular polygon (e.g., [Rodgers et al., 2010a]), *n*-gons (e.g., [Carroll et al., 2007]) or a particular irregular shape (e.g., [Griggs et al., 2004; Grünbaum, 1992]); the curves should be convex (e.g., [Mamakani et al., 2011]). However, it is not always possible to satisfy specific constraints. For instance, while a Venn diagram with three curves can be drawn with circles (Figure 3.8A), a Venn diagram with four curves cannot [Ruskey et al., 2006] and so, as proposed by Venn [1880], it is either drawn using ellipses (Figure 3.8B) or by adding a non-convex closed curve to a Venn diagram with three circles (Figure 3.8C). The latter is a general mechanism to draw Venn diagram with more than three curves [Venn, 1880] (see survey [Ruskey and Weston, 1997]).
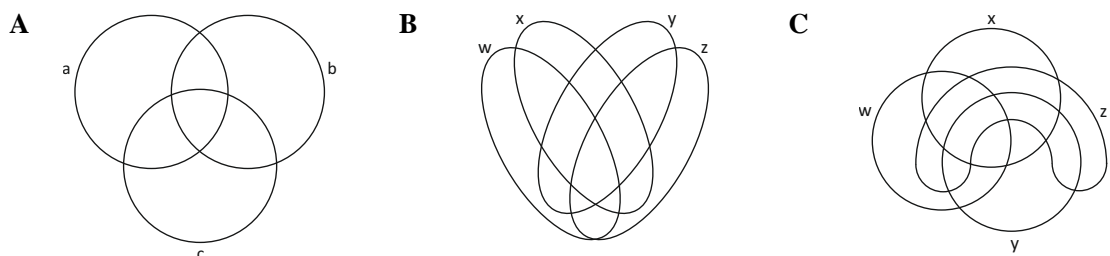
Figure 3.8: *A Venn diagram with three curves (left, A) and Venn diagrams with four curves (centre, B; right, C).* (A) A 3-Venn diagram drawn with circles. (B) A 4-Venn diagram drawn with ellipses. (C) A 4-Venn diagram drawn by adding a non-convex closed curve (curve *z*) to a 3-Venn diagram like A.

A diagram can have various other properties elicited by its curves. If a label is assigned to no more than one curve, the diagram has *unique curve labels*. Figure 3.9A has two curves labelled *c*, so it has duplicate curve labels. A diagram has *simple curves* when none of its curves self-intersect as curve *c* in Figure 3.9B. A diagram with *concurrent curves* has two or more curves sharing a common path for a segment of the curve. In Figure 3.9C, curves *a* and *b* are totally concurrent, as one is laid over the other,

and in Figure 3.9D, all the curves are partially concurrent to one another. A *brushing point* is when two curves do not cross each other but touch at one point, as curves *a* and *b* in Figure 3.9E. The point where *n* curves in a diagram meet is known as an *n-point*. Figure 3.9F has one 3-point or a triple point and three 2-points or double points. Figure 3.9B and E also have a triple point. A diagram has *multiple points* when it has one or more *n*-points with *n>2*. So Figure 3.9B-F all have multiple points.
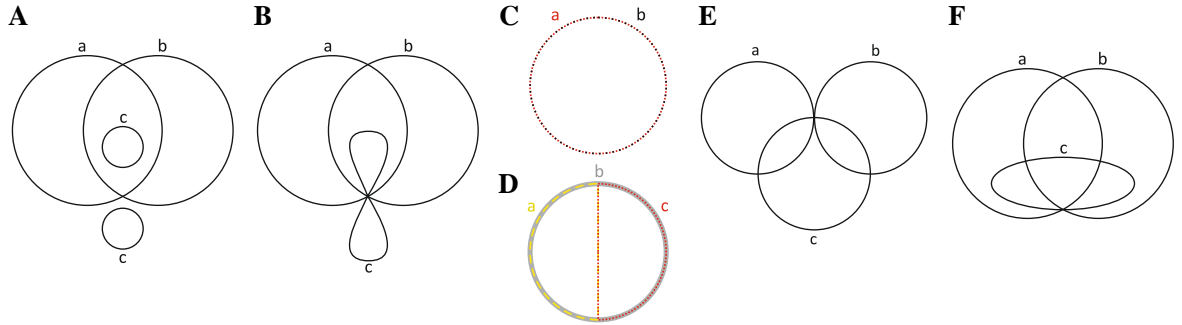
Figure 3.9: *Euler diagrams with (A) duplicate curve labels, (B) self-intersecting curves, (C) totally concurrent curves, (D) partially concurrent curves, (E) brushing points, and (F) multiple points.*

There might be constraints on the positioning of the curves. Example, symmetric Venn diagrams are *n*-Venn diagrams with *n*-fold rotational symmetry, so all the curves have the same shape and when the diagram is rotated $2\pi i/n$ for $i \in [0, 1, \dots , n\text{-}1]$ about its centre, it remains unchanged [Ruskey et al., 2006]. Figure 3.8A has 3-fold rotational symmetry and is a symmetric Venn diagram. Figure 3.8B-C are not symmetric and cannot be, as symmetric diagrams can only be drawn for a prime number of curves [Henderson, 1963]. Symmetric Venn diagrams have been studied (e.g., [Ruskey et al., 2006]), but it is unclear whether symmetry can aid diagram comprehension [Stapleton et al., 2007].

3.4.2 The Zones

A zone is described by the curves in which its minimal regions are located, so the *zone description* (or *zone label*) is the set of these curve labels. The zone with the minimal regions that are exterior to all of the curves in the diagram is labelled as ∅. Figure 3.10A indicates the zone labels (in red and italics) of the Euler diagram in Figure 3.1A. Example, the zone with the minimal region inside curves *a* and *c* is described and labelled as {*a,c*}, in short *ac*. The set of all the zone descriptions of a diagram is referred to as an *abstract description*. So, as shown in Figure 3.10A, the diagram in Figure 3.1A has zones ∅, *a*, *b, d, ab, ac* and abstract description {∅, *a, b, d, ab, ac*}. This example has connected zones, but the same applies to diagrams with disconnected zones, as shown in Figure 3.10B, which indicates the zone labels of the 3-Venn diagram in Figure 3.7B with abstract description {∅, *a, b, c, ab, bc, ac, abc*}.
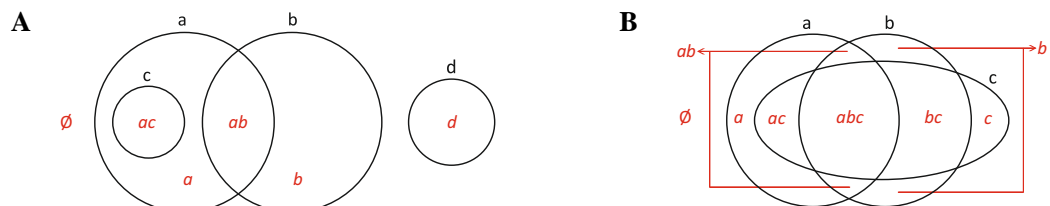
Figure 3.10: *Labelling the zones (in red and italics) of an Euler diagram with connected zones (left, A) and a Venn diagram with disconnected zones (right, B).* Curve labels are shown in black.

An abstract description is often defined before the diagram is drawn to describe the set of zones that are required. Automatic drawing algorithms for Euler diagrams typically take an abstract description as input and return an Euler diagram with exactly the set of zones described in the abstract description. If it is not possible to draw a diagram that represents the required zones and satisfies specified constraints like those in Section 3.4.1, the abstract description is said to be not drawable under those constraints.

A Venn diagram with n curves has 2^n zones (e.g., Figure 3.10B has three curves, thus eight zones). An Euler diagram with n curves can have one to 2^n zones (when it has 2^n zones, it is a Venn diagram). Given n curves, there is one set of zones that can be represented with a Venn diagram, but there are

$$\sum_{k=0}^{2^n}\left(\frac{2^n!}{k!\left(2^n-k\right)!}\right)$$

sets of zones that can be represented with an Euler diagram when all the possible combinations of various number of zones from the maximum 2^n zones are computed. So given three curves, there can be eight different zones and thus, one set of zones for 3-Venn diagrams and 256 sets of zones for 3-Euler diagrams. However, some of these sets of zones for Euler diagrams are isomorphic. For instance, the Euler diagrams in Figure 3.11 are the same up to curve label swapping. Stapleton et al. [2008a] define a number of invariants of isomorphic sets of zones and quote the number of non-isomorphic sets of zones that can be depicted with up to five curves (e.g., 80 non-isomorphic sets of zones for three curves). Sometimes, a diagram with a particular set of zones can be drawn in different ways using different aesthetic features (Section 3.4.1). For instance, all the diagrams in Figure 3.12 are 3-Venn diagrams.
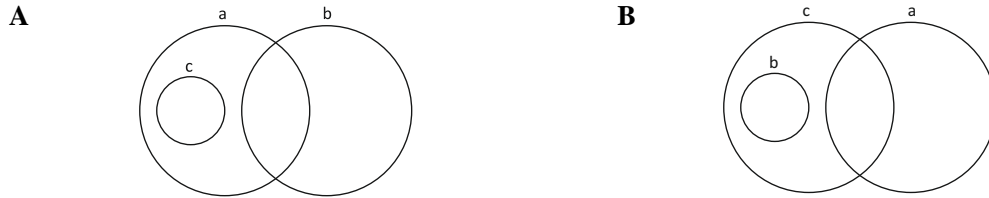
Figure 3.11: *An example of two isomorphic sets of zones represented by Euler diagrams.*
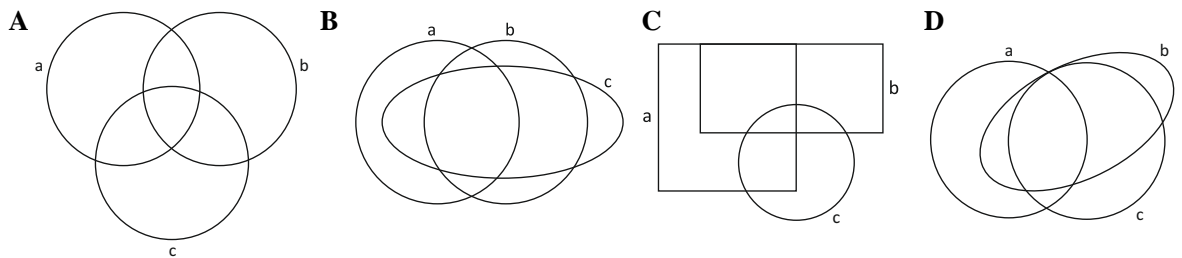
Figure 3.12: *Four possible ways how a Venn diagram with three curves can be drawn.* (A) All zones connected, no concurrent curves, no brushing points, no multiple points. (B) Two disconnected zones (*ab* and *b*). (C) All zones connected, but two concurrent curves (*a* and *b*). (D) All zones connected, but a brushing point (curves *b* and *c* touch at a point without crossing) and a triple point (the three curves meet at a point).

3.4.3 Representing Data Set Relations

Each data set is depicted by a curve in the diagram and each set relation is depicted by a zone. The label of each curve is the name of the data set it represents, so the zone label easily relates the data set relation it depicts and the abstract description clearly describes the data set relations the diagram is representing. A zone depicting a set relation that does not exist in the actual data is an *empty* zone.

If a Venn diagram is used, empty zones are shaded as proposed by Venn (Section 3.2, Figure 3.3). If an Euler diagram is used, all and only the zones corresponding to the required data set relations should ideally be visible. Venn diagrams become visually cluttered as the number of the curves increase, thus they are often used only when all the set combinations are required [Ruskey and Weston, 1997]. Venn diagrams may also be less effective than Euler diagrams for deductive reasoning [Sato et al., 2010].

Euler diagrams should not have empty zones, but this is not always possible if certain aesthetic features should be avoided. Example, in Figure 3.4C, the zone whose minimal region is in curves 'Ireland' and 'British Islands' but not 'United Kingdom' has no members and is thus empty. Rodgers et al. [2012b] argue that for an Euler diagram not to show this zone, either curves 'British Islands' and 'United Kingdom' are concurrent (Figure 3.13A) or the curve 'Ireland (island)' is non-simple (Figure 3.13B) eliciting a brushing point and a triple point. Alternatively two curves could be labelled 'Ireland (island)' [Rodgers et al., 2010b] (Figure 3.13C). Concurrent curves, non-simple curves, brushing points, triple points and duplicate curve labels impede diagram understanding and should be avoided (Section 3.5.1). Yet there are various sets of zones (e.g., those in Figure 3.9) for which a diagram with no empty zones cannot be drawn unless such undesirable features are used. In such cases, glyphs could be placed inside the non-empty zones, as in Figure 3.4C (the small blue ellipses with the names of the countries) and Figure 3.4B (the flags of the countries), or empty zones could be shaded (e.g., [Stapleton et al., 2012]).
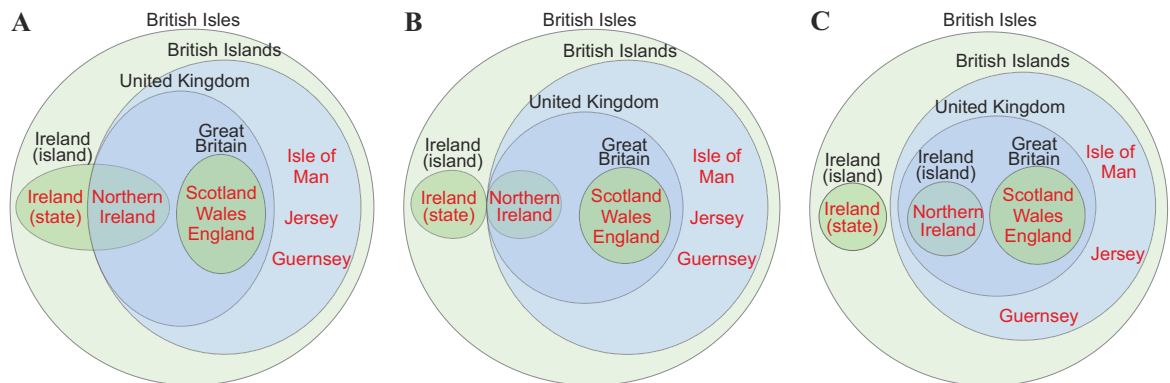
Figure 3.13: *The Euler diagrams proposed by Rodgers et al. [2012, 2010b] to represent the British Isles without showing empty zones (redrawn and adapted to match the data in Figure 3.4C).* Curve labels are shown in black just above the curve, while members are shown in red and italics inside the curves.

3.4.4 A Wellformed Diagram

A constraint that is often imposed when drawing an Euler diagram for an abstract description is that of wellformedness, in which case, the diagram must have the following *wellformedness properties*:

1. Connected zones (defined in Section 3.4);
2. Unique curve labels (defined in Section 3.4.1);
3. Simple curves (defined in Section 3.4.1);
4. Discrete curves, so the diagram has no concurrent curves (defined in Section 3.4.1);
5. Crossing curves, so the diagram has no brushing points (defined in Section 3.4.1);
6. Double points, so the diagram has no multiple points (defined in Section 3.4.1);

A *wellformed diagram* is one that satisfies all of these wellformedness properties. Figure 3.4C is a wellformed diagram, but Figure 3.13A-C are not. Figure 3.12A is also wellformed, but Figure 3.12B-D and all in Figure 3.9 are not. We discuss the importance of these properties in the next section.

3.5 Aesthetic and Perceptual Factors

The aesthetics of Venn and Euler diagrams have not been studied adequately, despite that the way they are drawn "will often influence how useful they are as tools" [Ruskey et al., 2006]. Perception "provides a sensible order to what we see, and aesthetics govern our receptiveness to our perceptions" [House et al., 2005, p. 87]. So their aesthetics should be considered with respect to human perceptual factors.

3.5.1 Wellformedness

The wellformedness properties in Section 3.4.4 were first noted by Flower and Howse [2002] in their paper describing the first drawing algorithm for Euler diagrams. These were later formalized [Stapleton et al., 2007] and used by various automatic algorithms (e.g., [Stapleton et al., 2008b, 2011c]). These properties are considered important in facilitating comprehension for the following reasons.

Not satisfying any of properties 1 and 2 (i.e., connected zones and unique curve labels) results in a dispersion of minimal regions for the same zone, thus an increase in cognitive effort to search, merge and track all of the minimal regions that make up a zone and a curve [Rodgers et al., 2012b; Sweller, 1994]. Not satisfying any of properties 3 to 6 (i.e., simple curves, discrete curves, crossing curves, double points) increases the amount of detail at specific points leading to additional cognitive effort to distinguish the curves as discrete complete objects [Rodgers et al., 2012b].

There are sets of zones for which an Euler diagram with precisely the required zones cannot be drawn without breaking one or more of these wellformedness properties (Section 3.4.3) and knowing which of these properties are most important could be helpful [Rodgers et al., 2012b]. In other cases, diagrams that are non-wellformed are generated despite that wellformed ones exist for the required set of zones (e.g., [Chow and Ruskey, 2004; Simonetto et al., 2009]). Studies assessing the effects of such properties on comprehension could thus provide guidance in the design of these diagram and automatic techniques that draw them. Only the following two studies of this kind have been conducted.

Rodgers et al.'s [2012b] study involved two experiments: one using diagrams with factual data (students registered for different modules) and another with abstract data (with curves labelled A, B, …) both of showing elements in the zones. Their second experiment suggests that wellformed diagrams improve comprehension. Both experiments indicate that: non-wellformed diagrams with disconnected zones or concurrent curves (breaking property 1 or 4) are the least helpful and most difficult to use; diagrams with non-simple curves (breaking property 3) are the least preferred; diagrams with brushing points (breaking property 5) do not have any negative effect on comprehension.

Contrary to these findings, Fish et al.'s [2011] study suggests that diagrams with concurrent curves (breaking property 4) can aid comprehension and diagrams with either brushing points or multiple points (breaking property 5 or 6) are the least helpful and the most difficult to use, but they confirm that wellformed diagrams improve comprehension. They also claim that diagrams breaking more than one of the wellformedness properties have a greater impact on comprehension than those breaking only one.

Discrepancies between the two studies could be due to a number of differences. In Fish et al.'s [2011] study, the tasks were oriented to logic and the diagrams had abstract data with no members depicted in the zones and with shading to indicate empty zones. Also, Fish et al. [2011] only tested restricted forms of diagrams with non-simple curves or concurrent curves (breaking property 3 or 4). The diagrams in the former had non-simple curves that self-intersected at one or two points and in the

latter the concurrent curves depicted total concurrency so that one curve was shown with two distinct labels. Thus, Fish et al.'s [2011] findings for properties 3 and 4 are not generalizable to all possible cases, particularly diagrams with partial concurrency. These inconsistent findings indicate the need for studies to assess these properties in various contexts using different design features and experimental parameters to ensure that findings are generalizable and representative of the class of such diagrams.

3.5.2 Curve Aesthetic

As mentioned earlier in Section 3.4.1, the curves in the diagram can have any shape, but some constraints could be imposed. This often depends on the context, the data being visualized, user preferences and aesthetics based on perceptual and cognitive factors that could aid comprehension.

A Specific Shape

The Euler diagrams drawn by Euler [1768a] and his predecessors Leibniz [Couturat, 1901] and Lull [1617] all used circles. Venn diagrams cannot always be drawn with circles [Ruskey and Weston, 1997] but when possible, Venn [1880] used circles or ellipses as in Figure 3.8. This indicates that curves that are simple and regular are often preferred. In fact, Birkhoff's [1933] aesthetic measure states that highly regular objects consisting of a few elements are likely to be seen attractive. After all, visual patterns that are mathematically simple are easier to recognize than random ones [Klinger and Salingaros, 2000]. Also, the structure and physical layout of visual representations affect the way readers perceive and understand the presented information [Ziemkiewicz and Kosara, 2010a; Ziemkiewicz, 2010].

Having curves with the same shape could be a rational choice. Shape is a preattentive feature [Ware, 2012, p. 154]. If a diagram has curves with a different shape from the majority, the reader's attention could be attracted to the different curves. Unless such a salience effect is explicitly required, it is often undesired, particularly in data analysis when each data set should be considered equally in an unbiased manner. Zones should also be similar in size (unless their area indicates some quantitative data), else the reader might think the size is encoding some information such as importance or quantity. Also, a study [Benoy and Rodgers, 2007] indicates that Euler diagrams with similarly sized zones aid comprehension.

The simple circle is widely used and preferred. An informal survey that analysed 72 Venn and Euler diagrams in Nature, Science and online affiliated journals in 2009 found that 65 of them (i.e., 90%) used circles [Wilkinson, 2012]. A circle is smooth, regular and recognizable, and could indicate "a sense of wholeness" [Ziemkiewicz and Kosara, 2010a] which contrasts the "regimented or compartmentalized structure" and isolation suggested by rectangles [Ziemkiewicz and Kosara, 2010a]. However, an Euler diagram with circles cannot be drawn for all sets of zones [Stapleton et al., 2012]. A wellformed Euler diagram can be drawn with circles for a class of abstract descriptions that have lots of set disjointness and subsets (i.e., 'Pierced' Euler diagrams) [Stapleton et al., 2011c], but for other abstract descriptions, the diagram with circles could have empty zones that are often shaded [Stapleton et al., 2012]. In such cases, circles and ellipses could first be used wherever possible and then arbitrary shaped curves, that are regular, smooth, convex as possible, could be added where required [Stapleton and Rodgers, 2011].

Continuity, Closure and Convexity

A study [Benoy and Rodgers, 2007] shows that smooth curves with good continuity aid comprehension of Euler diagrams. The Gestalt continuation principle [Koffka, 1935] also suggests that smooth curves

with good continuity are easier to identify than non-smooth ones. Studies [Field et al., 1993; Hess and Field, 1999] confirm that the ease of identifying a curve is a direct function of its good continuity: curves made up of connected small line segments whose orientation successively differs by ±30 degrees are more easily seen than when it differs by ±60 degrees. Also, curves with good continuity pop out as discrete, complete objects [Ware, 2012], so are easily distinguishable [Treisman and Gormican, 1988].

Figure 3.14A will be perceived as a rectangle and a circle (Figure 3.14B) rather than as two irregular shapes with abrupt changes in direction (Figure 3.14C). Though a rectangle is not smooth, a circle is and its good continuity facilitates identification of the two shapes in Figure 3.14A. Occam's razor plays an important role here as the simplest and most understandable explanations are chosen [Chater, 1997]. This also relates to the Law of Prägnanz [Koffka, 1935], which states that complex less recognizable objects are perceived as simpler, better understood elements. Yet it is unclear whether Figure 3.15A is two overlapping rectangles (Figure 3.15B) or three adjacent rectangles (Figure 3.15C). To imply the former, labels could be added to the curves (Figure 3.16A), but ideally the rectangles should be replaced with two smooth curves like circles (Figure 3.16B) so the two curves pop out as discrete objects. Also, Figure 3.16B is wellformed, but Figure 3.16A is not as it has concurrent curves. As noted in Figure 3.9, the curves in non-wellformed diagrams are not easily distinguishable as it is hard to follow the curves.
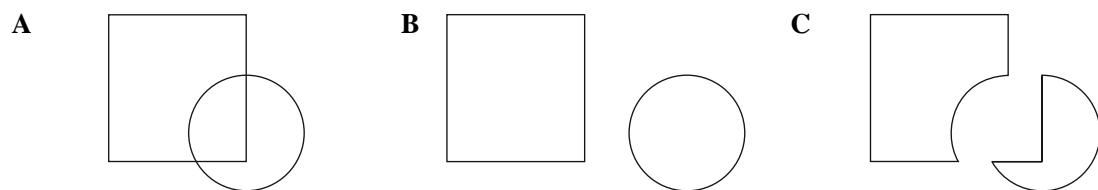
Figure 3.14: *Good continuity.* It is more likely for A to be interpreted as a rectangle and a circle (as in B) that are overlapping than two irregular shapes (as in C) that are touching.
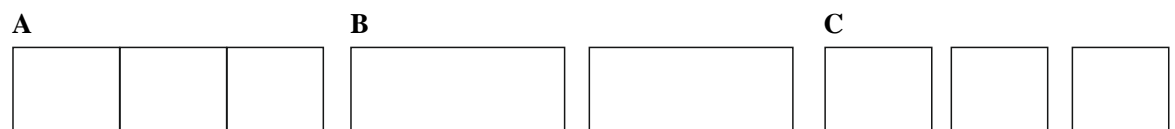
Figure 3.15: *Lack of good continuity.* It is unclear whether A is two rectangles (as in B) that are overlapping or three rectangles (as in C) that are touching.
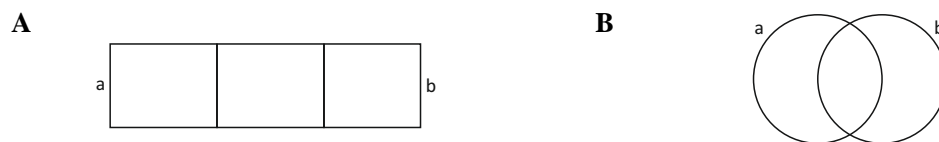
Figure 3.16: *Venn diagrams with two curves.* (A) The curves are rectangles. (B) The curves are circles.

Continuity has a stronger perceptual effect than colour and similarity and aids closure by facilitating the identification of closed curves even when they are partially occluded, but it is often underestimated [Ware et al., 2002]. Continuity and closure are crucial [Ware, 2012, p. 186] in supporting the principle of common regions [Palmer, 1992] and should thus be given priority. Convexity of the curves is also important as it too allows readers to perceive complete objects [Kanizsa and Gerbino, 1976] even when partially occluded [Liu et al., 1999]. Rectilinear curves could yet be used when the zone areas encode quantitative data as they could facilitate estimation and comparison of the quantities (Section 3.7.3).

Other Attributes

Curve closeness should also be considered. A study [Benoy and Rodgers, 2007] suggests that diagrams with segments of different curves close to one another impede understanding. In fact, Ware et al. [2002] argue that since neurons that detect orientation at the early stages of visual processing are approximately tuned within ±30 degrees [Blake and Holopigian, 1985], curves crossing at acute angles are more likely to lead to 'visual confusion' than ones that cross at a right angle. Though this was claimed in the context of graph drawing, this could be applicable to Euler diagrams and should be investigated further.

Every diagram design feature, be it the curve shape or any others that do not necessarily encode any specific information (e.g., colour schemes, outline thickness, background), should be considered with caution as they could affect the readers' diagram interpretation [Ziemkiewicz and Kosara, 2010b]. Also, some visual attributes that are meant to be meaningless could unintentionally put emphasis on specific aspects of the presented information [Elting et al., 1999]. Colour could be distracting, imply a semantic meaning, or attract attention especially when the coloured shape is large [Ziemkiewicz and Kosara, 2010a]. However, the design and perception of Euler diagrams have not be studied adequately.

3.5.3 Other Aesthetic Features

The cardinality of the data sets and their relations could be depicted by the area of the curves and the zones in the diagram. Such diagrams, referred to as area-proportional, are discussed in Section 3.7.

Other aesthetic features which should be considered include orientation, shape and colours. The importance of these features in Euler diagrams was recently noted [Blake et al., 2012b] and a study showed that diagram orientation has no significant effect on comprehension [Blake et al., 2012a]. Yet other related factors still need to be investigated. Example, can a change in orientation of a diagram that has already been processed affect recall and comprehension due to mental map preservation [Eades et al., 1991]? This is important for Euler diagram transformation [Fish, 2009; Rodgers et al., 2010b].

Visual clutter can make diagrams less usable [Rosenholtz et al., 2005]. Few metrics for clutter in Euler diagrams have been proposed [John et al., 2006; John, 2005] and a study showed consistency between the measured and user perceived clutter [John et al., 2006]. Euler diagram variants, ComED and DupED, try to reduce clutter [Riche and Dwyer, 2010] (Figure C.16A). ComED represents each set by multiple rectangular curves linked with hyperedges. DupED depicts each set by a rectangular curve, duplicates the elements belonging to multiple sets and connects these elements with hyperedges. A study revealed that ComED was more helpful for counting elements than for identifying set membership, while DupED significantly improved accuracy and the time taken for most of the tasks in the study [Riche and Dwyer, 2010]. However more work is required in the aesthetics of Euler diagrams.

3.5.4 Future Considerations

Aesthetic criteria, readability metrics and cognitive measures have to be defined, tested and formalized, as in example graph drawing [Dunne and Shneiderman, 2009; Purchase, 2002; Ware et al., 2002]. Drawing algorithms that optimize these criteria are more likely to generate readable diagrams. Aesthetic metrics for Euler diagrams that measure the roundness, smoothness and convexity of curves and zones, how distant curves are from one another, and how similar curves and zones are in size were defined, and two methods that optimize these metrics were devised to improve the aesthetics of already generated diagrams [Flower et al., 2003a]. Yet none of the metrics were fully formalized and tested.

Design guidelines based on perceptual theories such as Gestalt principles [Koffka, 1935] and Norman's [2004] emotional design model should be considered. Norman's model was originally intended for the evaluation of the commercial products, but it is applicable to information visualization. Norman [2002] makes a link between human emotions and cognition and states that attractive objects trick our receptors to view details in the object. Thus, effective visualizations are ones that engage the user and attract their attention [Healey and Enns, 2002; Tateosian et al., 2007], so that voluntarily they put in more effort in understanding the presented information. This concept was also confirmed in a study assessing the effect of aesthetic on information visualization [Cawthon and Moere, 2007]. Differences among users (e.g., spatial abilities, gender) should be taken into account as these can impact diagram understanding and usability [Ziemkiewicz and Kosara, 2009]. Thus studies comparing hand drawn and automatically generated diagrams should be conducted to aid in the identification of preferred visual features that users want to have and use (as in other visualization areas; e.g., [Alper et al., 2011; Dwyer et al., 2009]). A "user centered evaluation method not solely centered around task efficiency metrics is now imperative" [Cawthon and Moere, 2006] and so, available information visualization metrics that measure user experience [Cawthon and Moere, 2007] could be used.

Models of design in information visualization (e.g., [Liu and Stasko, 2010; Vande Moere and Purchase, 2011]) could aid in the understanding and evaluation of Euler diagram aesthetics. The graphical perception [Cleveland, 1994] of Euler diagrams should be studied to understand how various features (e.g., the shape or closeness of the curves) are perceived and to identify appropriate ways how to encode information in the diagram (e.g., Should the cardinality of a set relation be encoded by the area of the zone or by the number of glyphs in the zone? How is the area of a zone judged?). Besides the syntax of the diagram and how individual features are perceived and read, the overall interpretation of the diagram and its semantics should be investigated. This is important as simple aesthetic features and design choices (e.g., curve thickness, colours, background) can affect the mental model the reader constructs of the presented information [Ziemkiewicz and Kosara, 2010b].

Proposed formalisms and models of aesthetics [Filonik and Baur, 2009] could also be considered for Euler diagrams. Example, Birkhoff's [1933] aesthetic measure states that highly complex (made of a large number of elements) and disordered (made of a large number of irregularities) objects are unlikely to be seen attractive. The aesthetic score is defined as *order/complexity*, so the square gets the highest score out of various polygons as it is made of a few edges and has lots of symmetry [Birkhoff, 1933].

Aesthetics affect the usability and effectiveness of any information visualization [Cawthon and Moere, 2007] and thus, it should be given priority in the current ongoing research on Euler diagrams.

3.6 Automatic Drawing Techniques for Venn and Euler Diagrams

Drawing Venn and Euler diagrams is difficult particularly when a large number of curves and overlaps are required [Verroust and Viaud, 2004]. Automatic drawing techniques have been proposed. It is not always possible to draw an Euler diagram for any abstract description and satisfy specific aesthetic features (Section 3.5), so different techniques have been devised to meet different objectives.

It could be computationally expensive to explore the various ways how the required Euler diagram could be drawn, so often drawing algorithms provide an approximate or a heuristic solution [Chow, 2007]. Approximate solutions produce diagrams that, in some cases, either do not show all the required zones (e.g., [Kestler et al., 2008; Wilkinson, 2012]) or show empty zones which are often shaded (e.g.,

[Howse et al., 2005; Stapleton et al., 2012]) or noted by placing glyphs in all of the non-empty zones (e.g., [Clark, 2008]; the work in Section 6.3). Heuristic solutions vary from computational expensive ones such as evolutionary optimization (e.g., [Kestler et al., 2008]) to simpler and more efficient ones such as hill-climbing (e.g., [Chow and Rodgers, 2005; Flower et al., 2003b]; the work in Chapter 5).

We now discuss some of the main drawing techniques that are used. A detailed review is provided by Ruskey and Weston [1997] and Simonetto [2011] for respectively Venn and Euler diagrams.

3.6.1 Methods Using a Planar Dual Graph

A common practice adopted by a few Euler diagram drawing techniques (e.g., [Flower et al., 2008; Simonetto et al., 2009; Stapleton et al., 2011b]) is to create a planar dual graph of the zones in the abstract description and then use the graph to draw the curves of the diagram.

Figure 3.17A is a planar dual graph for abstract description $\{\varnothing, a, b, ab, bc, abc\}$. Each node in the graph corresponds to a required zone. Edges indicate zone adjacencies and where possible connect nodes whose label differs by only one curve label. This graph is embedded in the plane so the edges intersect only at endpoints. This elicits a planar dual graph (Figure 3.17A). The diagram is then drawn by enclosing all the nodes with a common curve label in one closed curve (Figure 3.17B). The resulting curves are often not smooth and easily distinguishable, so a few of these drawing techniques [Rodgers et al., 2008a, b; Simonetto et al., 2009] use a force-directed heuristic to improve the layout of the planar dual graph, before using it to draw the curves. In this way, Figure 3.17C is generated instead of the Euler diagram in Figure 3.17B [Rodgers et al., 2008b]. Yet the curves are often still non-smooth.
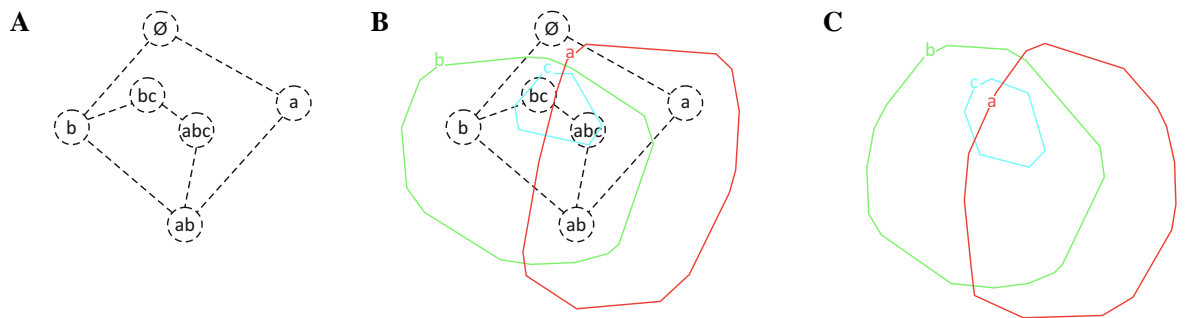
Figure 3.17: *Generating an Euler diagram for abstract description $\{\varnothing, a, b, ab, bc, abc\}$ using a planar dual graph.* The displayed graph and diagram were generated by Rodgers et al.'s [2008b] algorithm. (A) The planar dual graph (before its layout is improved). (B) The diagram generated after curves are drawn around the nodes with a common curve label. (C) The generated diagram when the layout of the planar dual graph in A is first improved by a force-directed heuristic. *Source:* [Rodgers et al., 2008b]—Figure 3 redrawn

3.6.2 General and Restricted Methods

Some drawing methods are general and draw an Euler diagram for any set of zones (e.g., [Simonetto et al., 2009; Stapleton et al., 2011a, 2011b]). Others are restricted to a class of abstract descriptions for which a diagram that satisfies some criteria can be drawn (e.g., [Chow, 2007; Stapleton et al., 2011c]).

The first automatic drawing algorithm that was devised [Flower and Howse, 2002] was restricted to the class of abstract descriptions for which a wellformed Euler diagram can be drawn. Other techniques with such a restriction were later proposed (e.g., [Rodgers et al., 2008b; Stapleton et al., 2011c]). Most

general drawing techniques try to draw wellformed diagrams whenever possible (e.g., [Rodgers et al., 2008a; Stapleton et al., 2010]), with a few also allowing users to specify their preferred wellformedness properties (e.g., [Stapleton et al., 2011b; Stapleton and Rodgers, 2011]). The generated diagrams could still be aesthetically undesirable, as various other aesthetic features should also be taken into account (Section 3.5). Thus, other techniques that restrict the shape (e.g., [Stapleton et al., 2011c, 2012]; Section 3.4.1) and form (e.g., [Mamakani et al., 2011]) of the curves were devised.

3.6.3 Layout Methods

Sometimes a different layout for an Euler diagram is required. A layout method transforms the current layout of the diagram into another that has the same set of zones, but satisfies specific aesthetic features.

A layout technique based on transformations of the Euler graph (i.e., a node is placed at each curve intersection point and undirected edges follow the direction of the underlying curve segment in the diagram) has been proposed to satisfy different wellformedness conditions [Rodgers et al., 2010b].

Another layout technique [Flower et al., 2003b] uses one of two implemented variants of a hill-climbing heuristic to optimize the weighted sum of a set of aesthetic metrics that the authors defined to generate improved layouts with curves that are round, smooth and not too close to one another and with curves and zones that are similarly sized and easily visible. According to a later study [Benoy and Rodgers, 2007], these features aid comprehension. This technique allows users to choose the criteria to be optimized and to provide a weight for each one. This could be helpful when multiple criteria cannot be fully optimized, as weights indicate how much a criteria should be enforced at the expense or benefit of another, but assigning appropriate weights is difficult [Marler and Arora, 2004; Rosenthal, 1985].

3.6.4 Inductive Methods and Drawing Diagrams From Existing Ones

An inductive drawing method [Stapleton et al., 2011b] is also available. Given an abstract description, curves are consecutively added to a new or existing diagram to get the required zones and to satisfy user preferred wellformedness properties. Similar methods were proposed by Venn [1880] and Edwards [1989] for Venn diagrams. Other techniques generate Euler diagrams from existing ones to preserve certain wellformedness properties (e.g., [Stapleton et al., 2008b]).

3.6.5 Methods for Euler Diagrams with Specific Features

Other algorithms were devised for Euler diagrams that are area-proportional (Section 3.7), 3D [Rodgers et al., 2012a] and those generated from hand drawn sketches [Delaney et al., 2010]. Various drawing methods have also been proposed specifically for Venn diagrams [Ruskey and Weston, 1997], despite that general methods for Euler diagrams can also draw Venn diagrams.

3.6.6 Methods for Euler-like Diagrams

A number of drawing techniques have been devised to enclose elements in Euler-like curves to visualize their set memberships and the set relations. Sometimes the elements have a predefined position and the curves are drawn on an existing visualization (e.g., Bubble Sets [Collins et al., 2009], Kelp diagrams [Dinkla et al., 2012]). If the position of the elements is not predefined, the elements are moved as the curves are laid out (e.g., [Simonetto et al., 2009], Set Visualiser [Wyatt et al., 2009; Wyatt, 2010],

ComED, DupED [Riche and Dwyer, 2010]). The curves are often rectangles (e.g., ComED, DupED), smooth (e.g., Bubble Sets, Set Visualiser) or non-smooth (e.g., [Simonetto et al., 2009]) irregular convex or concave curves, or curves with a schematic-like layout (e.g., Kelp diagrams). Sometimes the curves are not easily distinguishable due to example curve concurrency (e.g., [Simonetto et al., 2009]).

Other drawing techniques have been devised to depict clusters in graphs as a cartographic map (e.g., [Gansner et al., 2009, 2010a, b]) and to draw Voronoi diagrams [Aurenhammer, 1991]. However, the curves in these diagrams never overlap and are thus not applicable to Euler diagrams.

3.7 Area-Proportional Venn and Euler Diagrams

The zones in an Euler diagram represent specific data set relations and thus, the existence of elements that are members of the data sets in that relation. There are cases where the reader wants to see the actual elements, and glyphs or labels representing these elements are placed inside the zones. In other cases, the reader is interested in the total number of the elements in each zone. A label with the quantity could be placed in the zones but often, Euler diagrams that are area-proportional are preferred, so that the area of the zones and the curves reflects their cardinality, as in Figure 3.18A and Figure 3.18B. So differences in cardinalities between the data set relations and the data sets are easily noticeable.
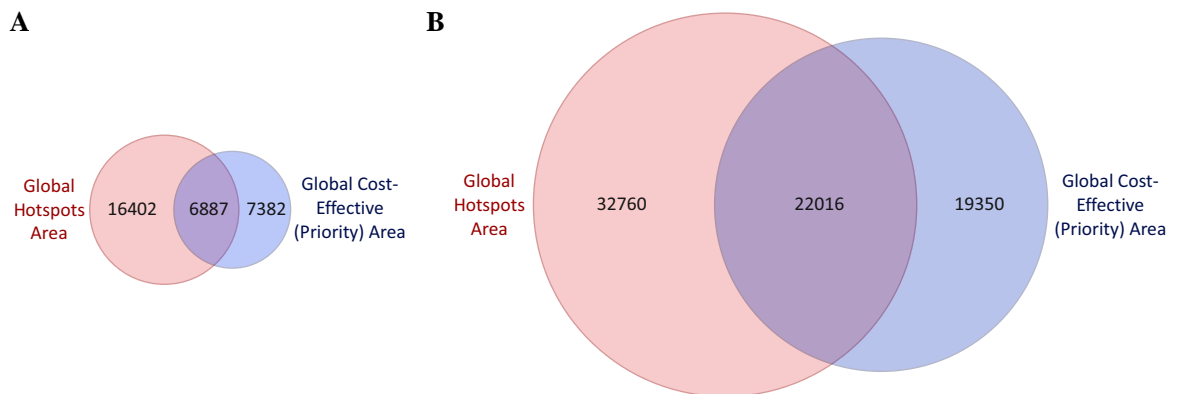
Figure 3.18: *Area-proportional 2-Venn diagrams showing the global overlap in land area (per 1000 km^2) between hotspots area and cost-effective priority area as determined by two global cost-effective conservation planning schemes, (A) the biodiversity hotspots and (B) the Global 200 ecoregions.* Data obtained from Table 1 of the manuscript, 'Global-scale mapping of economic benefits from agricultural lands: Implications for conservation priorities' [Naidoo and Iwamura, 2007]. The area of each region is directly proportional to the quantitative value displayed in the region. A common scale is used between the two diagrams so regions from different diagrams are comparable.

The notion of depicting quantity as area has long been used and Playfair seems to be the first [Tufte, 1983, p. 44]. For instance, Playfair [1801] encoded cities' population (Figure 2.2A) and countries' land mass (Figure 2.2B) by area of circles. Nightingale [1858] used the area of the radial segments in her iconic 'rose' diagram to report the number of deaths during the Crimean war (Figure 2.3). Studies show that humans are biased to area judgement, so area should not be used to encode quantitative data (Section 3.7.3). However, area-proportional Venn and Euler diagrams are widely used in diverse areas (Section 3.7.1) to facilitate data analysis.
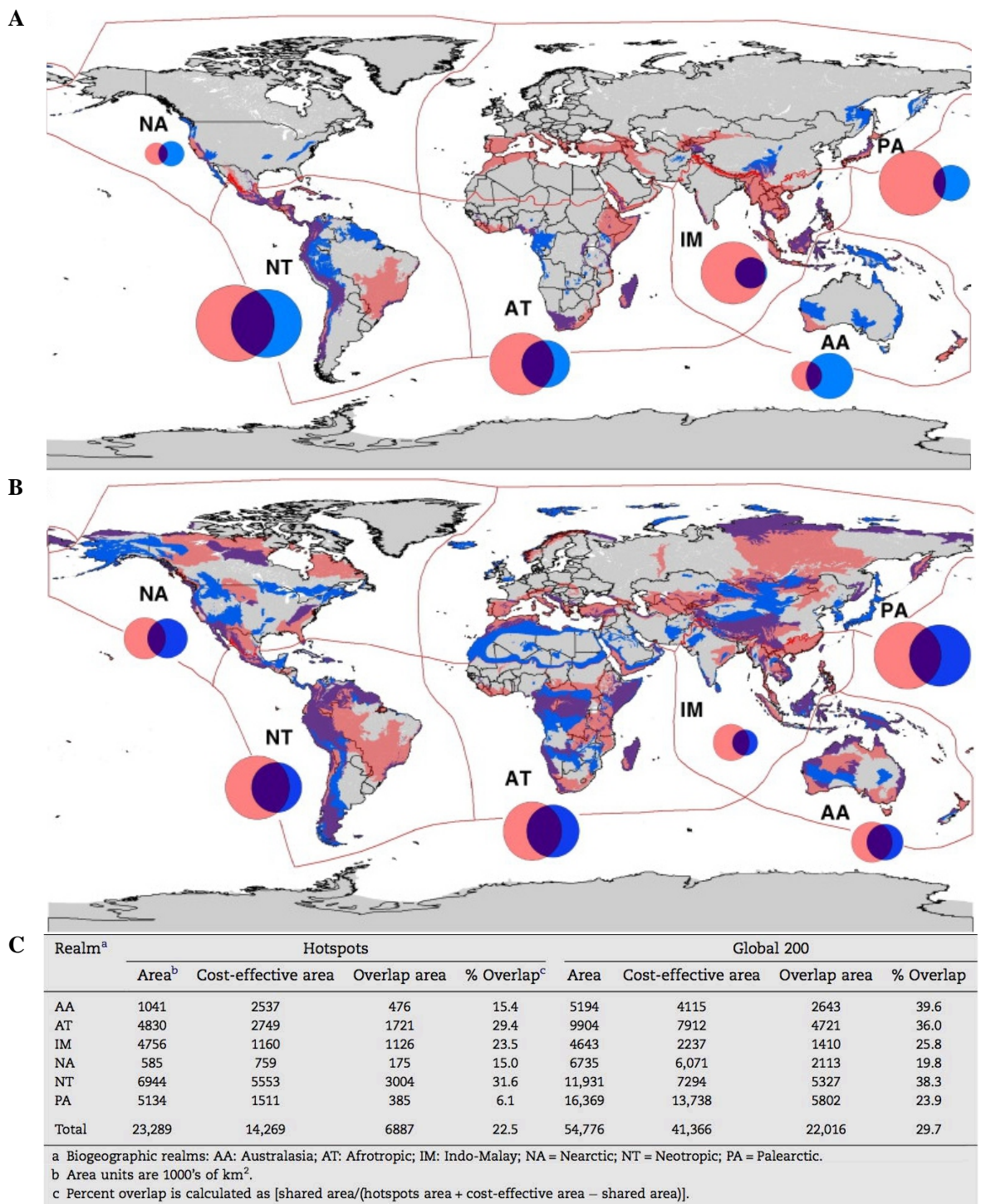
Figure 3.19: *Area-proportional 2-Venn diagrams showing the global overlap in land area (per 1000 km²) between hotspots area and cost-effective priority area as determined by two global cost-effective conservation planning schemes, (A) the biodiversity hotspots and (B) the Global 200 ecoregions, for six biogeographical realms to facilitate understanding of (C) the data in the table.* These are copies of Figure 4a, Figure 4b and Table 1 of the manuscript, 'Global-scale mapping of economic benefits from agricultural lands: Implications for conservation priorities' [Naidoo and Iwamura, 2007]. The hotspot areas are depicted in red and the cost-effective priority areas are depicted in blue. The six biogeographical realms include Nearctic (NA), Neotropical (NT), Afrotropical (AT), Indo-Malayan (IM), Australasian (AA), Palearctic (PA). The 2-Venn diagrams are depicting the quantitative data of the table in C.

Size is a preattentive feature [Treisman, 1985] that is easily noticeable and pops out from the rest of the display [Treisman and Gelade, 1980]. The pop-out effect of size is classified third just after colour and orientation which have a stronger effect [Ware, 2012, p. 155]. Size is also one of Bertin's visual or retinal variables [Bertin, 1983] which can easily be noted without much effort or cognitive processing. This could be one of the reasons why area-proportional Venn and Euler diagrams are widely used.

Area-proportional Venn and Euler diagrams also facilitate the comparison of quantitative data for the same sets but for a different factor (e.g., [Aronson et al., 2013; Clark et al., 2011; Nikulenkov et al., 2012]). For instance, Figure 3.18 illustrates the same data sets for two different schemes. Having a common scale and structure, similarities and differences are easily noticeable without the need to read the curve labels or to understand and process each of the diagrams. Tufte discusses similar benefits of using small multiples (i.e., same structured, small visual displays) to illustrate how a concept changes as a variable changes [Tufte, 1990, pp. 67-79]. In Figure 3.18, the small multiple is the 2-Venn diagram with data sets 'Global Hotspots Area' and 'Global Cost-Effective (Priority) Area', the concept is the percentage of the hotspot area that is prioritized and the variable between the diagrams is the scheme.

Hence, at first glance it is noticeable that the percentage of overlap between the two diagram is not so different, but the hotspot area in Figure 3.18B is clearly larger than that in Figure 3.18A. Though the diagrams in Figure 3.18 are not in the article [Naidoo and Iwamura, 2007], the authors take this concept further by drawing an area-proportional 2-Venn diagrams for each of the biogeographic realms for both schemes and displaying them on a map as in Figure 3.19A and Figure 3.19B to facilitate understanding of their tabular data in Figure 3.19C. Like this they demonstrate that the biodiversity hotspots conservation scheme "resulted in more priority areas in the low-opportunity-cost lands of the Nearctic and Australasia" and "in less area being prioritized in the Palearctic and Indo-Malay realms", while the Global 200 ecoregions scheme priortized least area in Nearctic, Palearctic and Indo-Malay which is visible from the small overlap [Naidoo and Iwamura, 2007, p. 44]. Using area-proportional diagrams, it is more likely to consider the overlap with respect to the size of the two curves and to view patterns in the data without looking at the actual values or curve labels. Avoiding labels by encoding the quantities as area could prevent adding noise to sophisticated diagrams like these [Ware, 2012, p. 170].

Venn diagrams with three curves drawn with circles are the most commonly used area-proportional Euler diagrams (Section 3.7.1). Yet drawing these diagrams accurately is difficult (Section 3.7.6). Thus, this dissertation focuses on area-proportional 3-Venn diagrams, unless notified otherwise. We now discuss application areas (Section 3.7.1), properties (Section 3.7.2), aesthetic and perceptual factors (Section 3.7.3), automatic drawing techniques (Section 3.7.4), theoretical findings (Section 3.7.5) and difficulties in drawing these area-proportional diagrams (Section 3.7.6).

3.7.1 Application Areas

Area-proportional Venn and Euler diagrams have long been used for various applications. Figure 3.20 and Figure 3.21 illustrate a few real-life examples. Other diagram examples in some medical disciplines include: medicine [Flowers et al., 2013]; genetics [Axtell, 2013]; vascular surgery [Ohrlander et al., 2011]; pathology [Short et al., 2013]; biosciences [Klees et al., 2007]; microbiology [Bielecki et al., 2013]; biochemistry [Aronson et al., 2013]; ecology [Liao et al., 2011]; neuroscience [Wang et al., 2010]; public health [Primack et al., 2013]. These diagrams are also used in non-medical disciplines like: astrophysics [Lister et al., 2013]; criminology [Farrell and Sousa, 2001]; segregation of duties [de Lange, 2011]; classification [Thièvre et al., 2005]; search and filtering [Dang et al., 2012].

An informal survey [Wilkinson, 2012] that analysed 72 Venn and Euler diagrams in 2009 Nature, Science and online affiliated journals found that 32 (44%) were area-proportional. Looking into these 32 area-proportional diagrams (after obtaining the list of the analysed articles from the author), we found that 14 (44%) were 2-Venn diagrams and 11 (34%) were 3-Venn diagrams (the remaining seven, that is 22%, were either Euler diagrams or Venn diagrams with more than three curves). All of these 2-Venn and 3-Venn diagrams were drawn with circles, as in Figure 3.20, except for a 3-Venn diagram drawn with two circles and an ellipse [Boj et al., 2009] and a 2-Venn diagram drawn with two squares [Hong et al., 2009]. This suggests that various real data sets are visualized using a 2-Venn or a 3-Venn diagram and that circles are preferred for these diagrams. The examples in Figure 3.21 are amongst the very few cases where polygons are used for 3-Venn diagrams, as often polygons and irregular curves are used for Venn diagrams with more curves (e.g., [D'Hont et al., 2012; Worden et al., 2009]).
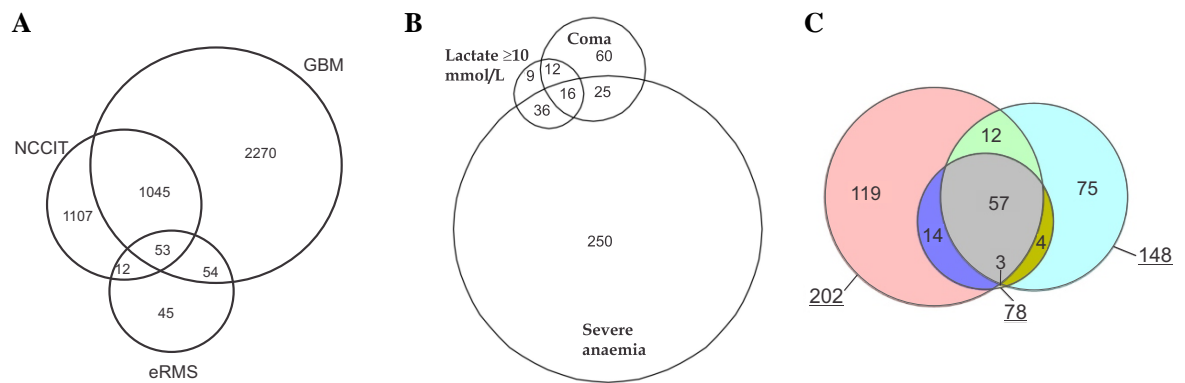
Figure 3.20: *Area-proportional 3-Venn diagrams drawn with circles in literature.* (A) Comparing the cell-type of differentially regulated genes after an anti-cancer drug treatment [Regenbrecht et al., 2008, p. 7]. The method used to draw this diagram was not noted. (B) Summarizing prognostic indicators of severe malaria [Dzeing-Ella et al., 2005, p. 4]. The method used to draw this diagram was not noted. (C) Analysing differences and similarities between identified chicken egg white proteome in three different studies [Mann and Mann, 2011, p. 3]. Drawn using Venn Diagram Plotter [Littlefield and Monroe, 2013].
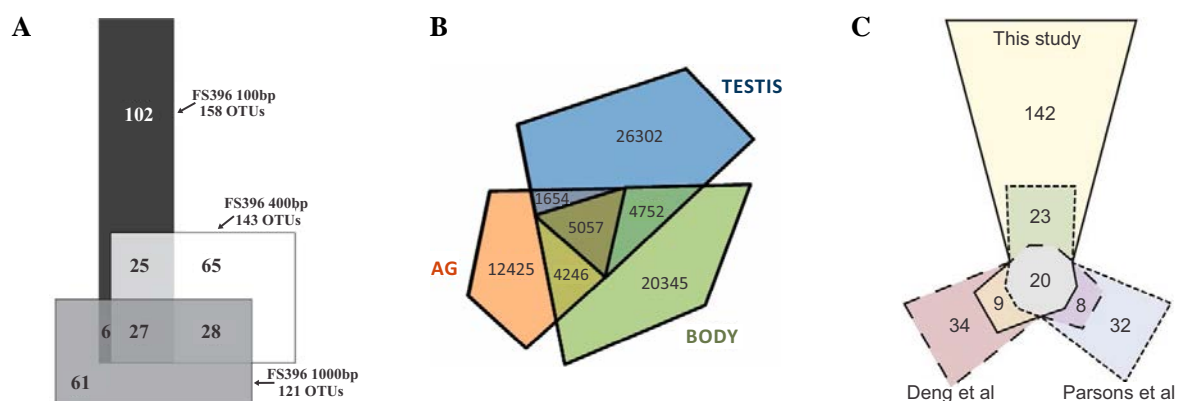
Figure 3.21: *Area-proportional 3-Venn diagrams drawn with polygons in literature.* (A) Analysing overlaps between gene libraries [Huber et al., 2009, p. 1297]. Drawn using DrawVenn [Chow and Ruskey, 2004]. (B) Studying transcriptome variation of different tissue types (i.e., testis, accessory gland, the rest of the body) of the male field cricket [Bailey et al., 2013, p. 227]. Drawn using Convex Venn-3 [Rodgers et al., 2010a]. (C) Summarizing genes affecting Top1-induced DNA damage identified in three different studies [Reid et al., 2011, p. 482]. Drawn using DrawEuler [Chow and Ruskey, 2005]. We added numeric labels to indicate the set intersection cardinalities (obtained from the article) that this diagram was intended to depict.

3.7.2 Properties

A Venn or Euler diagram is *area-proportional* when the area of each zone in the diagram is directly proportional to a quantity assigned to the depicted set relation. The required set of zones is indicated by the abstract description like not area-proportional diagrams (Section 3.4). The quantity assigned to each zone is indicated by an *area specification*, ω, so that if l is a zone label, $\omega(l) \in \mathbb{R}^+$ is the quantity assigned to the zone. The minimal regions of the zone labelled $\varnothing$ are exterior to all the curves, so one of these regions is unbound and infinite. Thus zone $\varnothing$ is not assigned a quantity and is excluded from the area specification. An example of an area specification defined for a Venn diagram with three curves labelled a, b and c is $\{(a, 141), (b, 139), (c, 179), (ab, 20), (ac, 40), (bc, 102), (abc, 39)\}$.

A zone is made up of one or minimal regions and thus the *area of a zone* is the sum of the area of all the minimal regions making up the zone. An area-proportional Euler diagram d is *accurate with respect to* an area specification ω, if d has all and only the required zones and for every zone z in d, the area of z, $area(z)$, is directly proportional to $\omega(z)$ and so $area(z) : \omega(z)$ is the same for all z. Else d is *inaccurate with respect to ω.* Direct proportions ensure graphical integrity, but other proportionality measures for accuracy could be defined and used (Section 3.7.3 subsection 'Counteracting Area Judgement Bias'). Figure 3.22 illustrates an example of an accurate and an inaccurate area-proportional Euler diagram (in this figure both examples are 3-Venn diagrams) with respect to specific area specifications. Formal definitions for area-proportional Euler diagrams are available [Chow and Ruskey, 2004; Chow, 2007].
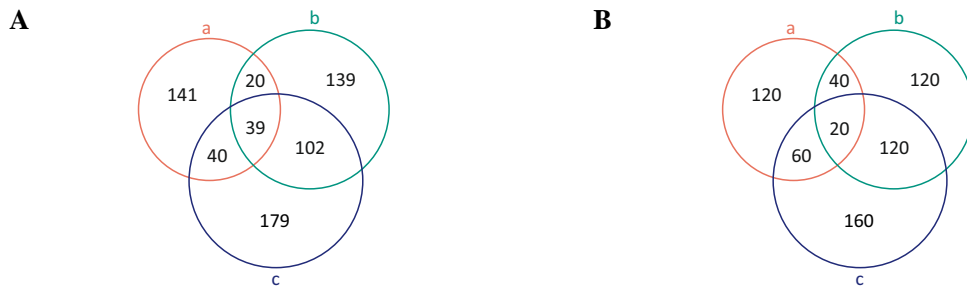
Figure 3.22: *An accurate (left, A) and an inaccurate (right, B) area-proportional 3-Venn diagram with respect to area specification* $\{(a, 141), (b, 139), (c, 179), (ab, 20), (ac, 40), (bc, 102), (abc, 39)\}$ *for A and* $\{(a, 120), (b, 120), (c, 160), (ab, 40), (ac, 60), (bc, 120), (abc, 20)\}$ *for B. The numeric label in each zone indicates the quantity the area of that zone should be proportional to, as specified in the area specification.*

3.7.3 Aesthetic and Perceptual Factors

The aesthetic and perceptual factors in Section 3.5 are also applicable to area-proportional Venn and Euler diagrams. Certain factors could in fact have a greater effect. Yet other factors, particularly those related to the perception of area, are more specific to area-proportional diagrams.

In the first paper on area-proportional Venn and Euler diagrams, Chow and Ruskey [2004] argue the importance for aesthetics that facilitate: (a) the identification of the curves in which zones are located; (b) the comparison of the zone areas. The former is discussed in Section 3.5, while the latter is discussed in this section. Yet the importance of different aesthetic and perceptual factors will also depend on how the diagrams are used in the application area. For instance, the curve areas could be judged and compared to obtain details on the size of the represented data sets.

Wellformedness

A diagram that breaks any of the wellformedness properties (Section 3.4.4) is difficult to comprehend (Section 3.5.1). If such a diagram is also area-proportional, the properties that make the diagram non-wellformed will have an additional adverse effect on the area perception of the zones and the curves.

If a zone in the diagram is disconnected (i.e., breaking wellformedness property 1), the zone is made up of multiple dispersed minimal regions. The area of a disconnected zone is the sum of the area of all of its minimal regions. Thus, increased cognitive effort is required to search for all the regions, estimate the area of each one and sum them to find the zone area. This means that comparing the area of different zones could be difficult if not impossible due to cognitive overload and a limited working memory [Sweller, 1994], which can handle only four elements [Cowan, 2000]. An increase in cognitive effort is also more likely to lead to errors and prolonged processing time [Kroger et al., 2002]. Same issues arise when a diagram has duplicate curve labels (i.e., breaking wellformedness property 2), as one data set would be depicted by the union of multiple curves. If a diagram breaks any of the other wellformedness properties, the curves are not easily distinguishable (Section 3.5.1), impeding the area perception of the zones and curves. Thus, this dissertation focuses on area-proportional diagrams that are wellformed.

The Shape of the Curves and the Zones

The shape of the curves and the zones in the diagram can have an effect on how their area is perceived. Cleveland argues that area judgement of regions on a geographical map is inaccurate and biased due to false misconception elicited by the shape of the regions. In Figure 3.23, Cleveland [1985, p. 282] notes that the state Florida (FL) seems bigger than Georgia (GA) despite the opposite is true, and Idaho (ID) seems bigger than Kansas (KS) despite their area is similar. Cleveland argues that FL, ID and OK seem bigger than their actual size due to appendages and their large perimeter. A non-smooth curve enclosing a region is likely to have a larger perimeter-to-area ratio than a smooth curve enclosing a region with the same area. Thus, smooth curves should be used when drawing area-proportional Euler diagrams as the region areas enclosed by the curves are less likely to be overestimated than with non-smooth curves.

Figure 3.23: *The map of U.S. states used by Cleveland [1985] to demonstrate the effect of shape on area perception. Source:* [Cleveland, 1985, p. 282]—Figure 4.38

However, rectilinear curves could be preferred to aid estimation and comparison of region areas. Example, in cartography, cartograms using rectilinear regions (e.g., [Buchin et al., 2012; de Berg et al., 2010]) are often preferred over those using arbitrary curves (e.g., [Dougenik, 1985; Sun, 2013]) to aid estimation of areas "by visual inspection" [de Berg et al., 2010, p. 203] (review [Speckmann, 2006]).

Chow and Ruskey [2004] argue that the area of zones in area-proportional Euler diagrams are easily comparable when the regions in the diagram are similar in shape, so two circles are easier to compare than a circle and a triangle, and rectangles are easier to compare when they have a similar aspect ratio. They state "the measure of a diagram's regional uniformity may be a good indicator of its effectiveness" [Chow and Ruskey, 2004, p. 476]. So according to the authors, Figure 3.24A could be effective as all of its regions are rectangles with an aspect ratio of $1:n$, $n \in \{1, 2, 4\}$, and a linear scaling. In Figure 3.24A, it is easily noticeable that zones a and b are twice as large as zones ab, abc and c, and four times as large as zones ac and bc, but this is not as apparent in Figure 3.24B despite that the diagram is accurate with respect to the same area specification of Figure 3.24A. Figure 3.24B is drawn with circles, so its regions have irregular shapes and a non-linear scaling. However, the curves in Figure 3.24A are not easily distinguishable and the zones are not easily identified, as in contrast to Figure 3.24B, the curves are irregular and non-smooth, and the diagram is non-wellformed (Section 3.5). So comparing the area of the curves in Figure 3.24B could be easier than in Figure 3.24A. This indicates that it is typically not possible to have easily comparable zones as well as easily comparable and distinguishable curves.
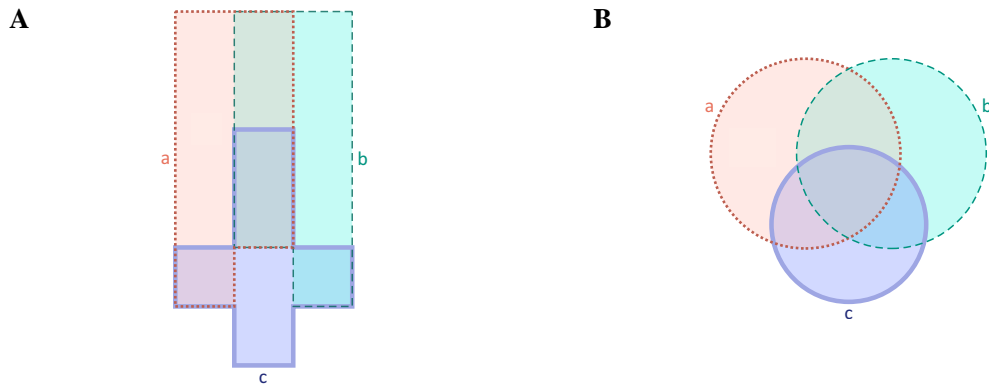
Figure 3.24: *Accurate area-proportional 3-Venn diagrams with respect to* $\{(a, 40), (b, 40), (c, 20), (ab, 20), (ac, 10), (bc, 10), (abc, 20)\}$ *drawn using rectilinear curves (left, A) and circles (right, B).*

Rectangular regions could aid area estimation, but comparing the area of rectangular regions that are squares or have extreme aspect ratios could be difficult [Kong et al., 2010]. The shape and position of the curves and regions in the diagram could also elicit optical illusions, like those in Figure 3.25, which can distort the perceived size [Coren and Girgus, 1978; Gregory, 1966]. Also "dimensions of shape that can be rapidly processed are unknown; however, the number is certainly small" [Ware, 2012, p. 171].
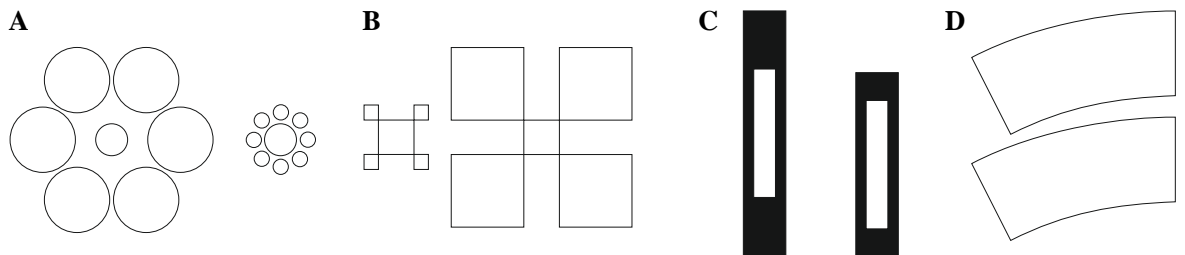
Figure 3.25: *Examples of optical illusions that could be relevant to area-proportional Euler diagrams.* (A) Ebbinghaus illusion. The circles at the centre of the 2 groups are congruent, but the left one (surrounded by larger circles) seems smaller than the right one (surrounded by smaller circles). (B) As A but with squares. (C) Both white rectangles are congruent, but the right one seems bigger as it has a thinner black border. (D) The shapes are congruent, but the bottom one seems larger. *Source:* [Stone, 2013]—figures redrawn

Counteracting Area Judgement Bias

Judging the area of the regions in Figure 3.24A might not be difficult, as they have a similar rectangular shape, the quantities assigned to the zones are $10n$, $n \in \{1, 2, 4\}$, and all the regions have the same width so it is enough to compare the height of the regions than estimate their area. Also, most regions have a common baseline, so the area of zones a, abc and b can be judged by comparing the position of their top edge with respect to the bottom and vice versa for zones a, ab and b and zones ac, c and bc.

The theory of visual perception and experiments in graphical perception [Cleveland and McGill, 1984, 1985; Cleveland, 1994] indicate that judging position along a common aligned scale is the most accurate task humans can carry out on graphical elements. Judging area is ranked fifth just after position along an identical but non-aligned scale, length and angles, and followed by volume and colour hue. Thus humans are biased to area judgement and the perceived area could be different from the actual area. This human limitation to perceive area has long been noted by Brinton in 1916 who claimed that

> Circles compared on a diameter basis mislead the reader by causing him to over-estimate the ratios. … Circles compared on an area basis mislead the reader by causing him to underestimate the ratios. Circles of different size should never be compared. Horizontal bars have all the advantages of circles with none of the disadvantages.
>
> [Brinton, 1916]

Stevens' [1957; 1986] Power Law on how scales are perceived states that, if readers are asked to estimate an area, $area_1$, with respect to another, $area_2$, most judgements would be of the form

$$\left(\frac{area_1}{area_2} \right)^{\beta}$$

where $\beta < 1$. Typically β is 0.7 [Stevens, 1975, p. 15]. So if $area_1$ is twice as much as $area_2$, readers are more likely to say $2^{0.7}$, that is 1.62 instead of 2, and if $area_1$ is five times as much as $area_2$, readers are more likely to say $5^{0.7}$, that is 3.09 instead of 5. It might sound sensible to use these laws to counteract the bias and "to take the areas to be proportional to the data to the $1/0.7 = 1.43$ power" [Cleveland, 1985, p. 282]. However, studies on area perception indicate that β could be any value between 0.6 and 0.9 [Baird, 1970, p. 64] as it depends greatly on the person and what is judging [Cleveland et al., 1982].

Other perceptual scaling measures have been proposed to scale symbols on maps and to compensate for the area judgement bias [Montello, 2002], but the great variations in the studies' findings indicate that area misperception is not systematic and it is unlikelihood that a universal 'correction' measure can be defined [Meihoefer, 1969, 1973]. The use of perceptual scaling was greatly criticized by Tufte who argued that irrespective of any possible perceptual limitations, it is compulsory "telling the truth about the data" [Tufte, 1983, p. 51] as this is the only way how "graphical excellence begins" [Tufte, 1983, p. 53]. In fact, his first principle to improve graphical integrity states that

> The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
>
> [Tufte, 1983, p. 56].

To help identify graphics that break this principle, Tufte [1983, p. 57] defined the 'Lie Factor' formula

$$Lie\ Factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data}$$

> If the Lie Factor is equal to one, then the graphic might be doing a reasonable job of accurately representing the underlying numbers. Lie Factor greater than 1.05 or less than 0.95 indicate substantial distortion.
>
> [Tufte, 1983, p. 57]

Experiments on area perception have been carried out only for circles [Cleveland and McGill, 1984; Flannery, 1971] and irregular regions [Cleveland and McGill, 1985; Cleveland, 1994] (squares for map symbols [Crawford, 1973]); none for Euler diagrams or areas of overlapping closed curves. Findings for non-overlapping objects might not be applicable to Euler diagrams, since they are highly dependent on what is judged [Cleveland et al., 1982]. No theories, metrics or studies provide guidelines on how area-proportional Euler diagrams should be drawn to aid judgement of the region areas. It is unclear which is the ideal shape for the regions, and the smallest perceivable region area and noticeable discrepancy in the region areas of such diagrams. Yet, even if these details are known, no effective and universal perceptual scaling measures could be defined. Thus area-proportional Euler diagrams should have zones whose area is "directly proportional to the numerical quantities represented" [Tufte, 1983, p. 56].

Different readers could judge the same area differently, so labels with the quantitative value depicted by each region could be added to the diagram (as in Figure 3.20, Figure 3.21) to reassure the perceived areas. Example, regions with the same area are easily noted in Figure 3.26 than in Figure 3.24 with the help of the labels. This supports Tufte's second principle to improve graphical integrity which states that

> Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself.
>
> [Tufte, 1983, p. 56]

Nonetheless, such an accurate judgement of the region areas is only required when looking into specific details of the data. If only an overview is required as in Figure 3.19, the labels showing the exact quantities should be omitted to avoid adding noise to the diagram [Ware, 2012, p. 170]. Instead of such labels, exact quantities can also be read from a table provided with the diagram, as in Figure 3.19C.
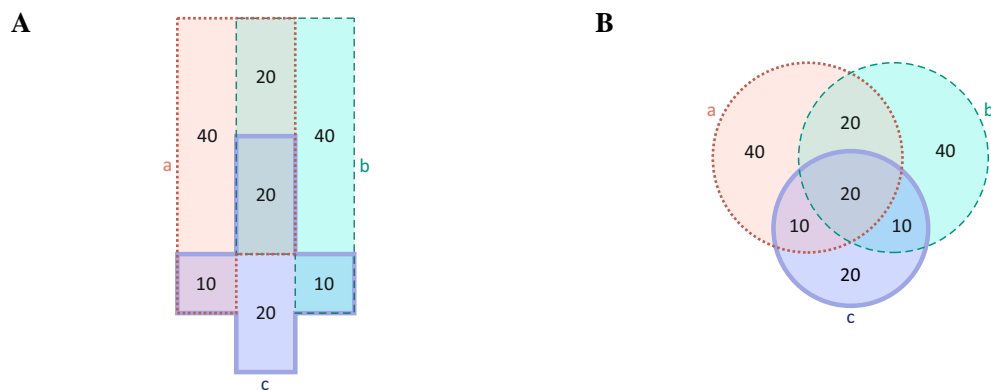
Figure 3.26: *Accurate area-proportional 3-Venn diagrams with respect to* {(*a*, 40), (*b*, 40), (*c*, 20), (*ab*, 20), (*ac*,10), (*bc*, 10), (*abc*, 20)} *with labels indicating the quantity each region represents.*

Cleveland argues that area judgement biases can be avoided only if area is not used altogether and quantitative data is instead encoded by position or length which humans can judge more accurately [Cleveland, 1985]. He explains that the population of European cities in Playfair's diagram with relatively sized circles (Figure 3.27A—Figure 2.2A replica) could be depicted more effectively using a dot chart with log scale (Figure 3.27B), so the position along a common scale rather than area is judged.

A

B
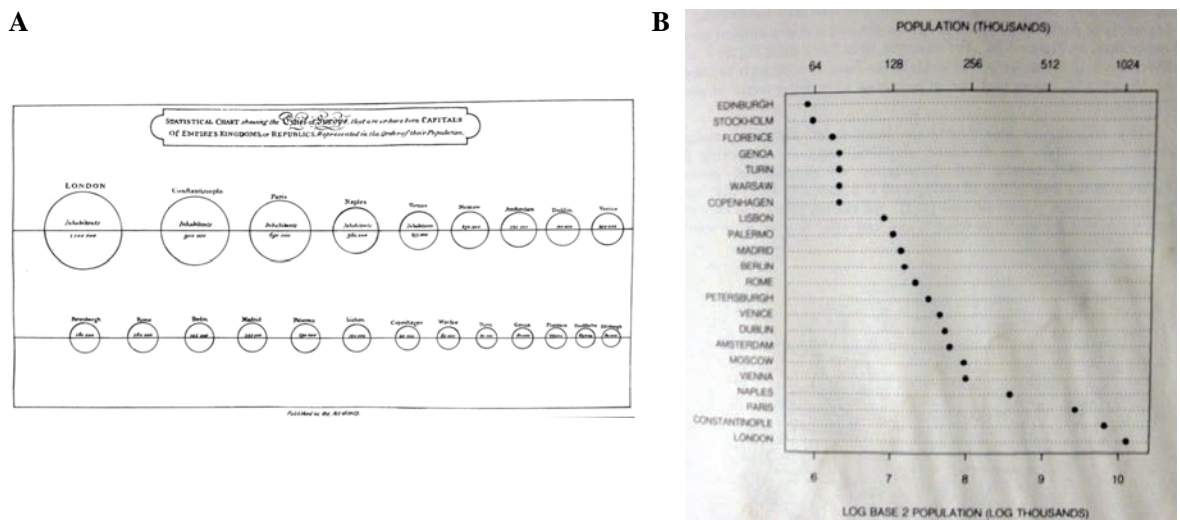
Figure 3.27: *Playfair's circles depicting the population of European cities (left, A) and Cleveland's dot chart with log scale for the same data (right, B). Source:* [Cleveland, 1985] (A) p. 115, (B) p. 281.

However, area-proportional Euler diagrams cannot easily be replaced by other diagrams. They are used extensively in various areas (Section 3.7.1), and in areas like genetics (e.g., [Bailey et al., 2013; el Azzouzi et al., 2011]), their use is often considered a standard and an essential part of the data analysis. Figure 3.24A attempts to divert area judgement to position and length judgement by representing the quantities as "columns of different height" [Bertin, 1983], but this is not always possible, particularly if the diagram has to be wellformed to aid understanding (Section 3.5.1). A dot or a bar chart encoding the quantitative data by position and length could be provided with an area-proportional Euler diagram. Alternatively to avoid redundancy, the quantitative data could be represented by the number of equally-sized glyphs in an Euler diagram that is not necessarily area-proportional. If the glyphs inside the regions are then placed in the form of a grid, the quantity could be judged by position or length judgement. Scaled glyphs could also be used. More on Euler diagrams with glyphs in Section 3.8.

3.7.4 Automatic Drawing Techniques

Current automatic techniques draw area-proportional Venn and Euler diagrams with circles or polygons.

Circle-based

The first automatic drawing methods to use circles were developed for area-proportional Venn diagrams with two [Chow and Ruskey, 2004] and three [Chow and Rodgers, 2005] curves. These were used in different areas, including medicine (e.g., [Lenz and Fornoni, 2006]). Later, other methods were devised (examples of diagrams in literature drawn using the method are given in parenthesis): BioVenn [Hulsen et al., 2008] (e.g., [Bouyer et al., 2011]); Venn Diagram Plotter [Littlefield and Monroe, 2013] (e.g., Figure 3.20C); a module in PatternLab for proteomics [Carvalho et al., 2008] (e.g., [Chun et al., 2011]); R packages, Vennerable [Swinton, 2009] (e.g., [Nikulenkov et al., 2012]) and `venneuler` [Wilkinson, 2012] (e.g., [Harvey et al., 2013]); Google Venn Charts [Google Inc, 2012]; Stata's PVENN [Gong and Ostermann, 2011]; SAS macro [Shiqun, 2009] (e.g., [Soriano et al., 2003]); Matlab's VENN [Darik, 2011] and vennX [Heil, 2004]; a web application [BioInfoRx Inc , 2013] (e.g., [Kang et al., 2011]).

Except for `venneuler`, all of these drawing methods aim to draw area-proportional Venn diagrams with two or three circles. Some allow a few zones to be assigned a value of zero (e.g., Vennerable, BioVenn) and thus try to draw Euler diagrams. The first proposed drawing method [Chow and Ruskey, 2004] draws accurate 2-Venn diagrams for any set of zone areas, so most later techniques used this method within their algorithm to draw 3-Venn diagrams. A few (e.g., BioVenn, Venn Diagram Plotter) are actually variants of the first drawing technique for 3-Venn diagrams [Chow and Rodgers, 2005].

The technique `venneuler` [Wilkinson, 2012] is particularly different from the rest. Using a statistical model for fitting an area-proportional diagram to an area specification comprising of a normalized loss function *stress* (defined as, the sum of squared residuals divided by the total sum of squares) and a mechanism to minimize the function, `venneuler` is considered to be the current best method in drawing accurate area-proportional Venn and Euler diagrams for any number of curves with circles. However, circles cannot draw accurate area-proportional 3-Venn diagrams for most data (Section 3.7.6) and thus, it still generates inaccurate and misleading diagrams for most data, as shown in example Figure 3.28.

In Figure 3.28A, zone *abc* with assigned value 1 appears to have the same area as zones *ab*, *ac* and *bc*, each of which should represent value 50 and all of which seem smaller than half the area of zones *a*, *b* and *c* that should depict 100 each. Figure 3.28B has missing zone *abc* and Figure 3.28C has missing zone *ac*, so both diagrams do not depict all the required set relations. Yet they both have a low *stress* of 3.12×10^{-4} and 5.85×10^{-3} respectively as indicators to `venneuler` that the diagrams are accurate. In some cases, the generated diagram is non-wellformed as in Figure 3.28C where zone *b* is disconnected. Similar problems are also apparent in other circle-based techniques. There are area specifications for which an Euler diagram cannot be drawn with circles unless unwanted zones are visible (Section 3.5.2). To handle this, circle-based methods for not area-proportional Euler diagrams shade unwanted zones [Stapleton et al., 2012] or use some non-circle shaped curves [Stapleton and Rodgers, 2011]. `venneuler` cannot shade or use non-circle curves, so some of its generated Euler diagrams have unwanted zones like *ac* and *bc* in Figure 3.28D, which can only be avoided if curve *c* is non-simple as in Figure 3.9B.
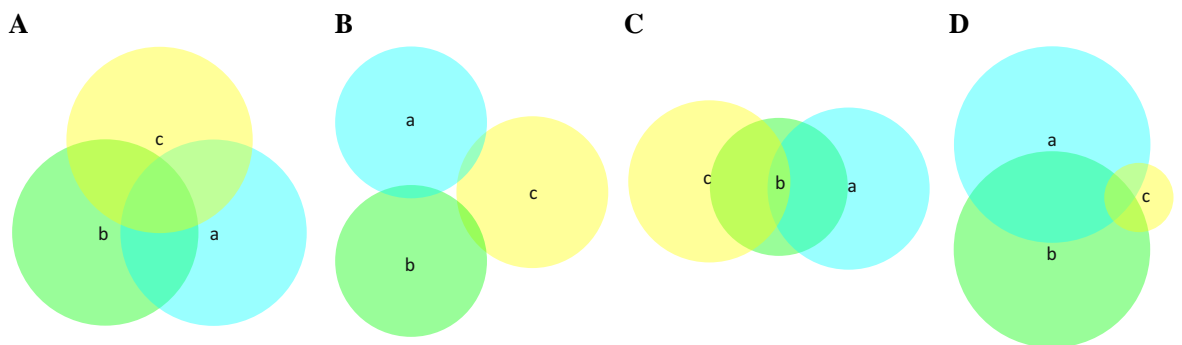
Figure 3.28: *Examples of area-proportional 3-Venn diagrams drawn using venneuler.* All of these diagrams are inaccurate with respect to the required area specification. (A) The generated diagram with respect to {*a*=100, *b*=100, *c*=100, *ab*=50, *ac*=50, *bc*=50, *abc*=1}, whose *stress* value is 4.77×10^{-2}. All the required zones are present, but their area is misleading with respect to the quantities they represent. (B) The generated diagram with respect to {*a*=100, *b*=100, *c*=100, *ab*=1, *ac*=1, *bc*=1, *abc*=1}, whose *stress* value is 3.12×10^{-4}. Zone *abc* is missing. (C) The generated diagram with respect to {*a*=100, *b*=10, *c*=100, *ab*=50, *ac*=10, *bc*=50, *abc*=10}, whose *stress* value is 5.85×10^{-3}. Zone *ac* is missing and zone *b* is disconnected. (D) The generated diagram with respect to {*a*=10, *b*=10, *c*=1, *ab*=5, *ac*=0, *bc*=0, *abc*=1}, whose *stress* value is 2.70×10^{-3}. Zones *ac* and *bc* are visible even though they were not meant to be included in the diagram.

As we discuss in Section 3.7.6, no circle-based method can draw accurate area-proportional Venn diagrams with three curves for most sets of zone areas due to the limited degrees of freedom of circles. Inaccurate diagrams can be misleading, so a few drawing methods use not area-proportional diagrams and either label the zones with the required value, as in R packages VennDiagram [Chen and Boutros, 2011] and Limma [Smyth, 2005], GeneSpring® [Agilent Technologies Inc, 2012], GeneVenn [Pirooznia et al., 2007], ArrayTrack [Fang et al., 2009], VENNY [Oliveros, 2007] and Pangloss Venn Diagram Generator [Seidel, 2005], or fill up the zones with equally-sized glyphs that are proportional in number to the value assigned to the zone, as in TwitterVenn [Clark, 2008]. VennPlex [Cai et al., 2013] uses not area-proportional diagrams with labels to report multiple values corresponding to each zone.

Algorithms that generate circular cartograms (scaled regions are depicted as circles) are available (e.g., [Dorling, 1996]), but are not applicable to Euler diagrams, as in cartograms no overlaps are shown.

Polygon-based

The first proposed method VennMaster [Kestler et al., 2005, 2008] uses convex regular polygons. Such polygons are similar in shape to circles, so the generated diagrams are often inaccurate (e.g., [Rodrigues et al., 2012]). Other methods use: triangles as in Vennerable [Swinton, 2009]; rectangles [Marshall, 2005]; orthogonal rectilinear curves as in DrawVenn [Chow and Ruskey, 2004] (e.g., Figure 3.21A) and Vennerable [Swinton, 2009]; 4-sided and 5-sided convex polygons as in Convex Venn-3 [Rodgers et al., 2010a] (e.g., Figure 3.21B); parallelograms [Wieland et al., 2005]; orthogonal polyominoes [Chow and Ruskey, 2007]; a combination of convex and non-convex, smooth and rectilinear curves as in VENNTURE [Martin et al., 2012] and in Vennerable [Swinton, 2009] (both using Edward's [1989] general n-Venn diagram construction); convoluted polygons as in DrawEuler [Chow and Ruskey, 2005] (e.g., Figure 3.21C) and Fan Diagrams [Kim et al., 2007]. A method that draws diagrams with polygons for any number of curves has been proposed but not implemented [Stapleton et al., 2011a]. A recent method, Euler3, was devised to use polygons only when circles cannot be used [Rodgers et al., 2014]. Often polygons are used for area-proportional diagrams that cannot be drawn accurately with circles.

Algorithms that generate cartograms using rectilinear (e.g., [van Kreveld and Speckmann, 2007]) or arbitrary polygon (e.g., [Sun, 2013]) curves are available, but are not applicable to Euler diagrams as in cartograms no overlaps are shown and some criteria to be optimized are different (e.g., a cartogram should preserve the region adjacencies of a geographical map). Algorithms for voronoi treemaps are also available (e.g., [Sud et al., 2010]), yet again no overlaps are shown as their curves do not intersect.

3.7.5 Theoretical Findings

Chow's [2007] theoretical findings on area-proportional Euler diagrams are the basis for understanding how and when these diagrams can be drawn accurately. The most important for this dissertation are:

1. An accurate area-proportional 2-Venn diagram with respect to any area specification can be drawn using circles ([Chow, 2007]—Theorem 3.1.1).

2. There are area specifications for which an accurate area-proportional 3-Venn diagram cannot be drawn using circles ([Chow, 2007]—Theorem 3.3.1).

3. There are area specifications for which an accurate area-proportional wellformed 3-Venn diagram cannot be drawn using convex curves ([Chow, 2007]—Theorem 3.4.2; [Rodgers et al., 2009]).

Other theoretical findings that are relevant to this dissertation include:

4. An accurate area-proportional wellformed 3-Venn diagram with respect to any area specification can be drawn using orthogonal rectangles or orthogonal 6-gons ([Chow, 2007]—Theorem 3.2.1).

5. An accurate area-proportional 3-Euler diagram with respect to any area specification can be drawn using orthogonal rectangles or orthogonal 6-gons ([Chow, 2007]—Theorem 3.2.3).

Additional work was carried out (e.g., [Stapleton et al., 2011a]), but is not relevant to this dissertation.

3.7.6 Difficulties in Drawing Accurate 3-Venn Diagrams

An accurate area-proportional Venn diagram can be drawn with circles for any quantitative data with two sets but not three (Section 3.7.5). Section 3.7.1 showed the diagrams in Figure 3.29, as examples of the 3-Venn diagrams with circles used in various application areas. All of these diagrams are inaccurate with respect to the quantitative data indicated by the numeric labels in the zones. In Figure 3.29A, the zone with value 45 is much bigger than those with values 53 and 54. In Figure 3.29B, the zone with value 25 is greater than that with value 36, and in Figure 3.29C, the zone with value 3 is barely visible and much smaller than that with value 4. Similar problems are apparent with most area-proportional 3-Venn diagrams drawn with circles, including those generated by `venneuler` [Wilkinson, 2012], which is considered the most effective in accurately drawing theses diagrams (Section 3.7.4). This is due to the limited degrees of freedom of circles (i.e., a centre and a radius). Thus, numeric labels are often added to the diagrams to confirm the quantitative data that the diagram is meant to represent.
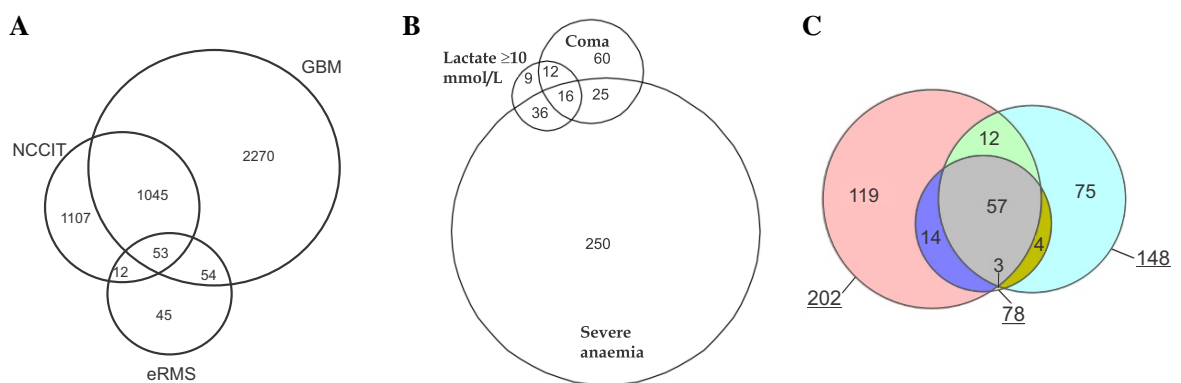
Figure 3.29: *Area-proportional 3-Venn diagrams drawn with circles in literature.* Figure 3.20 replica

In contrast to circles, an accurate area-proportional 3-Venn diagram can be drawn with polygons for any quantitative data (Section 3.7.5). Section 3.7.1 showed the diagrams in Figure 3.30, as examples of the 3-Venn diagrams with polygons used in various application areas. All of these diagrams are accurate with respect to the quantitative data indicated by the numeric labels in the zones. This is so as polygons have various degrees of freedom that can be adjusted to accurately depict the required data. Yet the irregularity and non-smoothness of polygons make the diagrams difficult to comprehend as the curves and their overlaps are not easily identified (Section 3.5.2, Section 3.7.3). This is particularly problematic when the curves are concurrent as in Figure 3.30A, and when the diagram has brushing or multiple points and the curves meet or cross at bending points as in Figure 3.30B and Figure 3.30C. The curves in Figure 3.30A and B are convex, so they are more likely to be perceived as complete objects [Kanizsa and Gerbino, 1976] than those in Figure 3.30C , but their pop-out effect is not as strong as with circles.
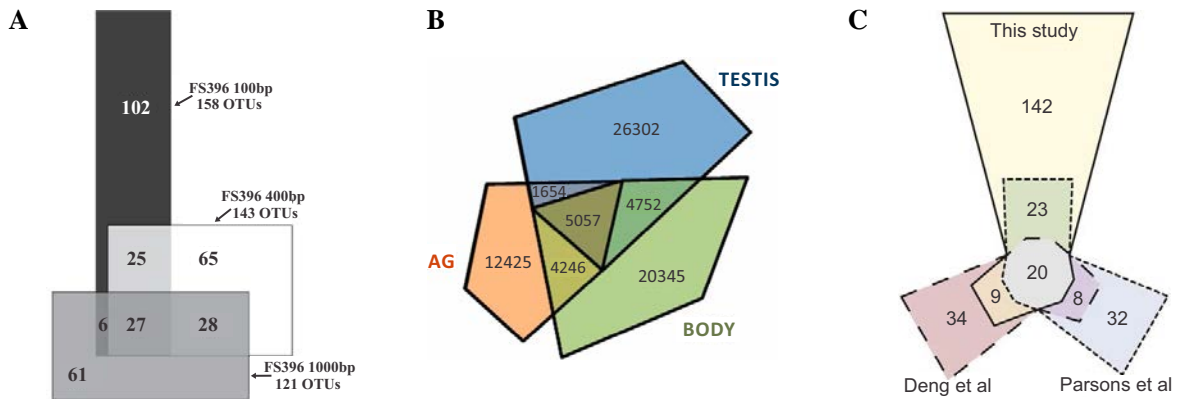
Figure 3.30: *Area-proportional 3-Venn diagrams drawn with polygons in literature.* Figure 3.21 replica

Hence, circles and polygons cannot draw area-proportional Venn diagrams that are accurate as well as easy to comprehend for most 3-set data. Yet drawing methods, prior to the work in Chapter 5, use circles or polygons. Irrespective of the curve shape, an accurate area-proportional Venn diagram cannot be defined analytically for a given area specification [Chow, 2007] and thus a heuristic is required.

The Need for a Heuristic

No analytic method can determine the properties of the curves of an accurate area-proportional Venn diagram for a given area specification, even when the diagram has as little as two curves and when simple shapes with very few properties (like circles) are used to represent the curves [Chow, 2007].

Consider for instance two overlapping circles. The area of their overlap is defined by the properties of the two circles, namely their radius and their centre. The radius of each circle can be computed if the area of the circles is known, but not the coordinates of their centre. Given a specific area for the overlap, the value of the properties of the two intersecting circles cannot be determined analytically as the function defining the overlapping area is non-invertible (details given by Chow [2007, pp. 61-64]). So numerical methods and heuristic search techniques (Section 3.7.4) are used to adjust the properties of the curves to generate an appropriate area-proportional diagram.

This problem is also evident with other curves that are more general than circles. Consider ellipses. The overlapping area of two ellipses is defined by their properties, namely the semi-minor and semi-major axes, the centre and the angle of rotation of each ellipse. It is more complex with ellipses than circles, as the area of an ellipse is defined by a function of the semi-minor and semi-major axes that is also non-invertible. So given the area of an ellipse, its two semi-axes cannot be determined analytically, unless the semi-axes are equal to one another in that the ellipse is a circle. Thus, an optimization technique is required to find an appropriate set of values for the properties of the intersecting ellipses and generate a diagram whose curve and zone areas accurately depict the required quantitative data.

Hence, irrespective whether the curves are depicted as circles, ellipses, polygons or any other shape, the function defining the zone areas is non-invertible [Chow, 2007]. We now explain how an accurate area-proportional 2-Venn can always be drawn with circles and why this is not possible for 3-Venn.

Constructing Accurate Area-proportional Venn Diagrams with Two But Not Three Circles

An accurate 2-Venn diagram as Figure 3.31A can be drawn for any area specification using two circles a and b (Section 3.7.5). This is fully constrained, as given that circles a and b have radii r_a and r_b respectively and distance d_{ab} between their centres as in Figure 3.31B, only one overlapping area exists.

So to draw the accurate 2-Venn diagram in Figure 3.31A, radii r_a and r_b and distance d_{ab} must be computed. The area of circles a and b is the sum of the required area of the zones inside each of the circles. Knowing the area of the circles, radii r_a and r_b are then computed using simple Euclidean geometry rules. If a is located at the origin such that its centre is $(0, 0)$, then the centre of b is $(d_{ab}, 0)$. We know that the value of d_{ab} is between $(r_a—r_b)$ and (r_a+r_b). If $d_{ab}= (r_a—r_b)$, the area of the overlap would be equal to the area of the smaller circle, as in Figure 3.31C, so that the smaller circle would be fully contained in the larger circle. As the value of d_{ab} increases, the area of the overlap decreases and is reduced to zero as in Figure 3.31D when $d_{ab}= (r_a+r_b)$. Thus, a numerical approximation technique such as the bisection method [Burden and Faires, 2010] can be used to find d_{ab} in the interval (r_a-r_b, r_a+r_b) as explained by Chow and Ruskey [2004] (also [Chow, 2007, pp. 61-64]). The diagrams in Figure 3.31C and Figure 3.31D are Euler diagrams not Venn diagrams as they fail to represent all possible set intersections and so, d_{ab} should not be equal to $(r_a—r_b)$ or (r_a+r_b). Yet, if Euler diagrams with two circles are required, the bisection method should be applied in the interval $[r_a—r_b, r_a+r_b]$.
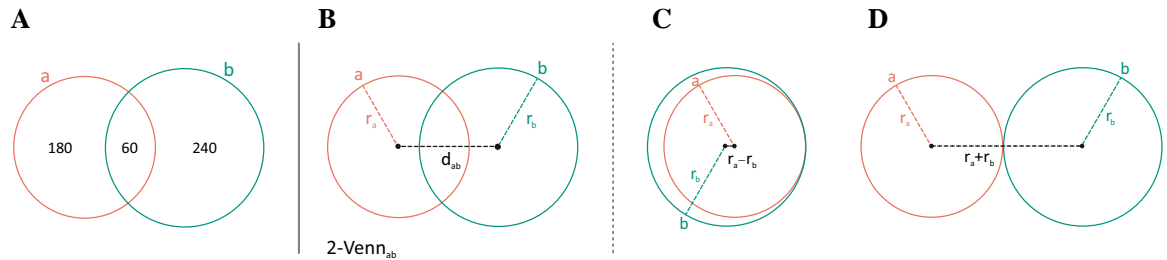
Figure 3.31: *Constructing an area-proportional 2-Venn diagram using circles and the bisection method.* (A) An accurate area-proportional 2-Venn diagram. (B) A 2-Venn diagram, *2-Venn*$_{ab}$, drawn with two circles a and b—the area of the overlap is defined by radii r_a and r_b and the distance d_{ab} between the centres of a and b. The value of d_{ab} should be (C) greater than $(r_a – r_b)$ but (D) smaller than (r_a+r_b).

When a circle c is added (Figure 3.32A), the overlapping region of the 2-Venn diagram is split up into two and new regions are introduced. So as shown in Figure 3.32B, a 3-Venn diagram is made up of three 2-Venn diagrams. Thus, an area-proportional 3-Venn diagram for the zone areas depicted by the numeric labels in Figure 3.33A (i.e., area specification {a=120, b=120, c=160, ab=40, ac=60, bc=120, abc=20}) can be constructed by first drawing the exact 2-Venn diagrams in Figure 3.33B, whereby two copies of circle c are drawn to ensure that both *2-Venn*$_{ac}$ and *2-Venn*$_{bc}$ are accurate. Then, rotate the left copy of c anticlockwise about the centre of a and the right copy of c clockwise about the centre of b. Only one circle c is required and so the two copies of c are rotated until they overlap (Figure 3.33C). By satisfying this constraint, the overlapping regions in a 3-Venn diagram are formed (Figure 3.33D). Yet the resulting zone areas (Figure 3.33E) are unlikely to coincide with those required (Figure 3.33A) and no changes can be made to get the required areas without making other areas inaccurate.
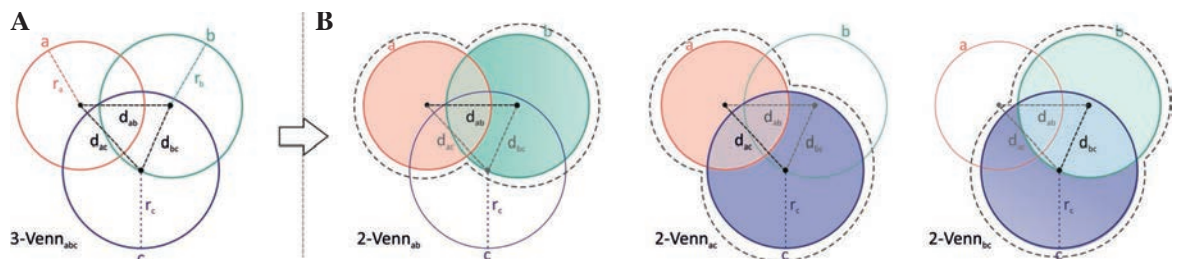
Figure 3.32: *A 3-Venn diagram is made up of three 2-Venn diagrams.* (A) A 3-Venn diagram, *3-Venn*$_{abc}$. (B) The three 2-Venn diagrams in *3-Venn*$_{abc}$: *2-Venn*$_{ab}$, *2-Venn*$_{ac}$, *2-Venn*$_{bc}$.
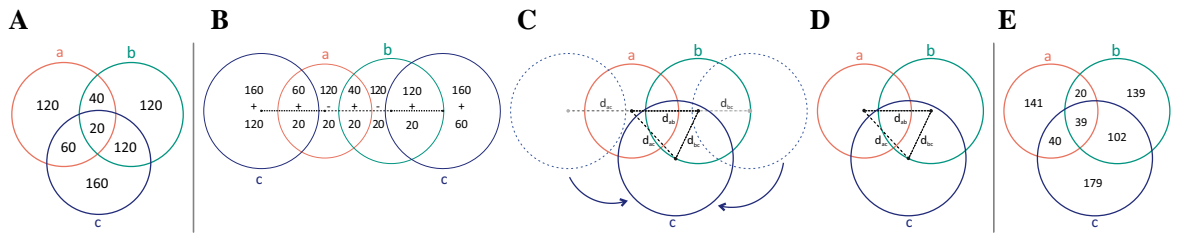
Figure 3.33: *A method for constructing an area-proportional 3-Venn diagram with circles.* (A) The numeric labels in the zones indicate the required zone areas, thus the area specification (i.e., {*a*=120, *b*=120, *c*=160, *ab*=40, *ac*=60, *bc*=120, *abc*=20}) for which an area-proportional 3-Venn diagram should be drawn. (B) 2-Venn diagrams with respect to {*a*=200, *c*=280, *ac*=80}, {*a*=220, *b*=280, *ab*=60} and {*b*=200, *c*=220, *bc*=140} respectively are drawn. (C) The two copies of circle *c* are rotated so they overlap and unite. (D) When only three circles are visible, a 3-Venn diagram is obtained. (E) The zone areas in D are not proportional to the desired values in A. The numeric labels in the zones indicate the zones' actual area.

If, instead of circles, polygons were used, an accurate area-proportional 3-Venn diagram could have been generated. Figure 3.34 shows such a diagram drawn with rectilinear curves (Figure 3.34A), convex 5-sided curves (Figure 3.34B), and complex non-convex curves (Figure 3.34C) for the quantities shown by the numeric labels in Figure 3.33A. However, as discussed earlier, these diagrams are difficult to use and comprehend as the curves are non-smooth and not easily identified. Instead of area, the quantitative data could be depicted in an Euler diagram using glyphs, as we discuss in the next section.
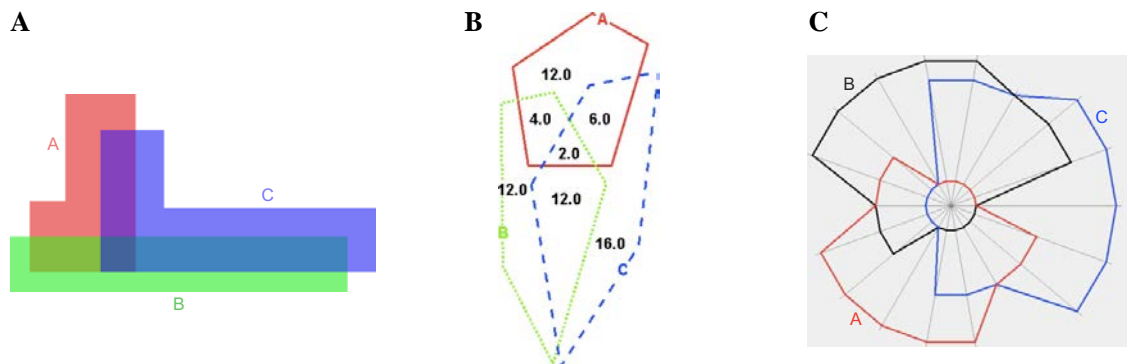
Figure 3.34: *Accurate area-proportional 3-Venn diagrams drawn with polygons with respect to area specification {a=120, b=120, c=160, ab=40, ac=60, bc=120, abc=20}.* Curve *a* is drawn in red, curve *b* in green (or black in the case of C), curve *c* in blue. These were generated using: (A) DrawVenn [Chow and Ruskey, 2004]; (B) Convex Venn-3 [Rodgers et al., 2010a]; (C) DrawEuler [Chow and Ruskey, 2005].

3.8 Euler Diagrams with Glyphs

A possible way to counteract the area judgement bias in humans (Section 3.7.3) is to place equally-sized glyphs that are proportional in number to the desired quantities in the regions of an Euler diagram that is not necessarily area-proportional. In this way, the quantity assigned to each zone in an area specification is represented by the number of glyphs in the regions of that zone in the Euler diagram and the data sets and their relationships are represented by the curves in the Euler diagram. This alternative to area was proposed by Bertin [1983] to represent size and quantitative data, example on maps.

Only few automatic methods draw Euler diagrams with glyphs. TwitterVenn [Clark, 2008] draws Venn diagrams with equally-sized glyphs that are proportional in number to the twitter messages that

used any of two or three user selected words (Figure 3.35). A zone with no glyphs is empty, thus glyphs can be used instead of Venn's [1880] shading mechanism for empty zones (e.g., Figure 3.3). Other methods draw irregular curves around elements forming Euler-like diagrams with glyphs (e.g., [Collins et al., 2009; Simonetto et al., 2009]; Section 3.6.6; examples in Appendix C.4). These methods differ from TwitterVenn as the glyphs also show the individual identity of the elements besides the quantity.
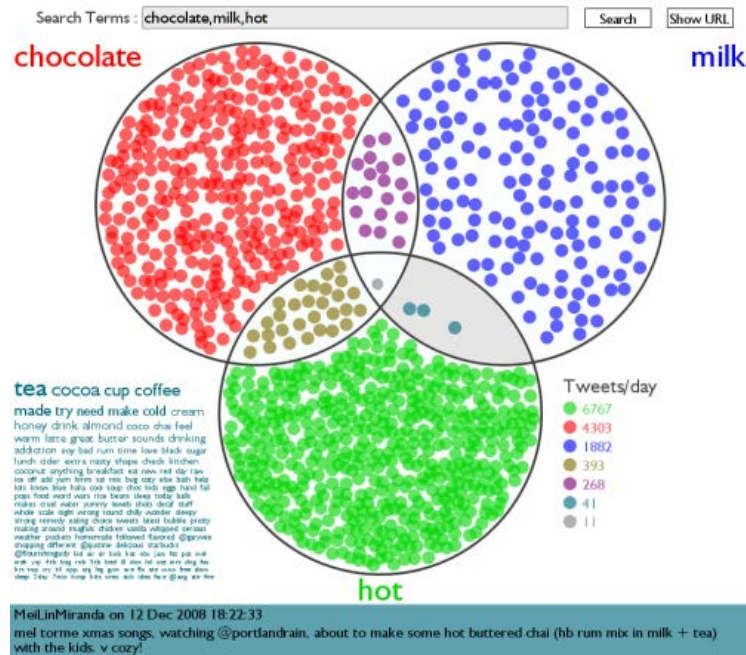
Figure 3.35: *A not area-proportional Venn diagram with equally-sized glyphs drawn by TwitterVenn.* The user selected words are 'chocolate', 'milk' and 'hot'. *Source*: [Clark, 2008].

Euler diagrams with multi-attribute glyphs have been proposed [Brath, 2012] (Figure 3.36) but not evaluated. Euler diagrams with glyphs have also been drawn manually for 2×2×2 contingency tables [Myšičková and Vingron, 2012] as well as Bayesian problems as we discuss in the next section.
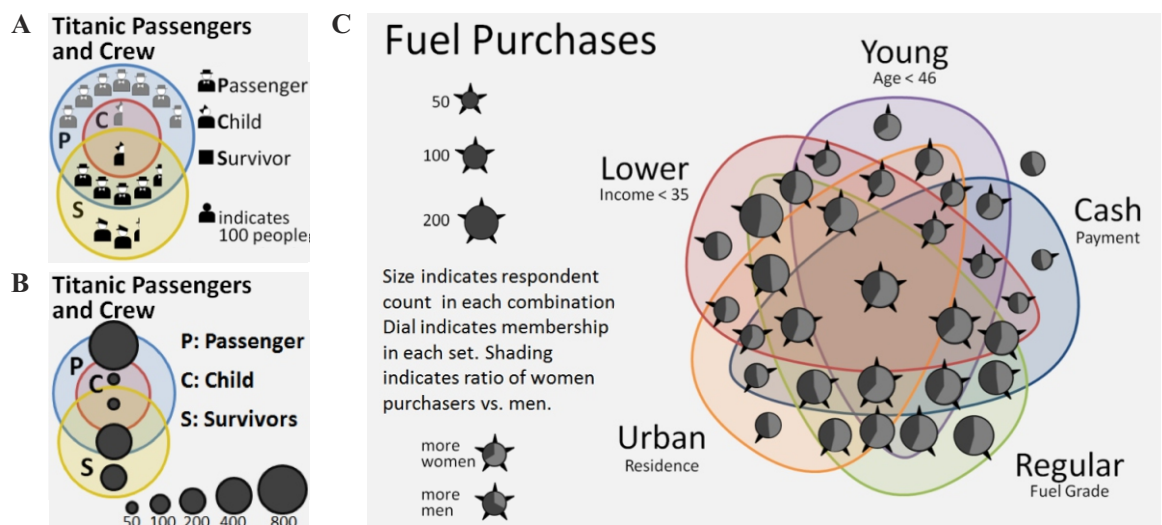
Figure 3.36: *A few of Brath's [2012] Euler diagrams with glyphs.* Quantities assigned to the zones are depicted by the glyphs as follows: (A) the number of pictograms; (B) the area of the scaled glyphs; (D) the area of whisker glyphs whose whiskers indicate the sets in which they are located and its glyph shading indicates the proportion of purchases by females versus males. *Source*: [Brath, 2012] Figure 4, 5, 10.

3.9 Euler Diagrams for Bayesian Reasoning

Making decisions based on statistical, uncertain data is difficult and often involves Bayesian inference, which is subject to fallacious reasoning (Section 2.5). Psychology studies assessed a few visualizations for Bayesian reasoning, but their effectiveness is still unclear. Euler diagrams and frequency grids, as in Figure 2.9, are amongst those evaluated. Studies on frequency grids were discussed in Section 2.5.5.

We now discuss how area-proportional Euler diagrams and Euler diagrams with glyphs have been used for Bayesian reasoning and what the studies on such visual representations revealed.

3.9.1 Area-Proportional Euler Diagrams

Sloman et al. [2003] argued that area-proportional Euler diagrams, as Figure 2.9A, are effective in conveying the nested-set relations of Bayesian problems (success of 48% of 25 subjects with Euler diagrams and 20% of 25 with text alone, over a revision of Casscells et al.'s [1978] disease diagnosis Bayesian problem with probabilities, among university students aged 18 to 24 who had completed courses in Psychology, Cognitive Science, Economics). However, similar benefits were reported when the nested-set relations were explicated in the text. Later Brase [2009] found no improvement with area-proportional Euler diagrams (success of 34.7% of 98 subjects with Euler diagrams and 35.4% of 96 with text alone, over a problem similar to Sloman et al's [2003] but with natural frequencies, among university Psychology students aged 18.8 on average, who participated to fulfil a class requirement).

Kellen [2012] (also [Kellen et al., 2013]) claimed improved performance when an area-proportional Euler diagram as Figure 3.37 was provided (correct means of 5.38 for 40 subjects with Euler diagrams and 3.32 for 37 subjects with text alone). However, he argued that this representation is too complex as there are various interacting elements. Instead he proposed Figure 3.38 as a relationally and cognitively simpler alternative and reported improved performance particularly when a diagram was not provided (correct means of 6.39 for 44 subjects with Euler diagrams and 7.41 of 37 with text alone). The results of this study also indicated that people with low spatial abilities have difficulties in solving such problems using a visual representation as, in contrast to those with high spatial abilities, they performed better with text alone than with a diagram (be it complex as in Figure 3.37 or simple as in Figure 3.38).
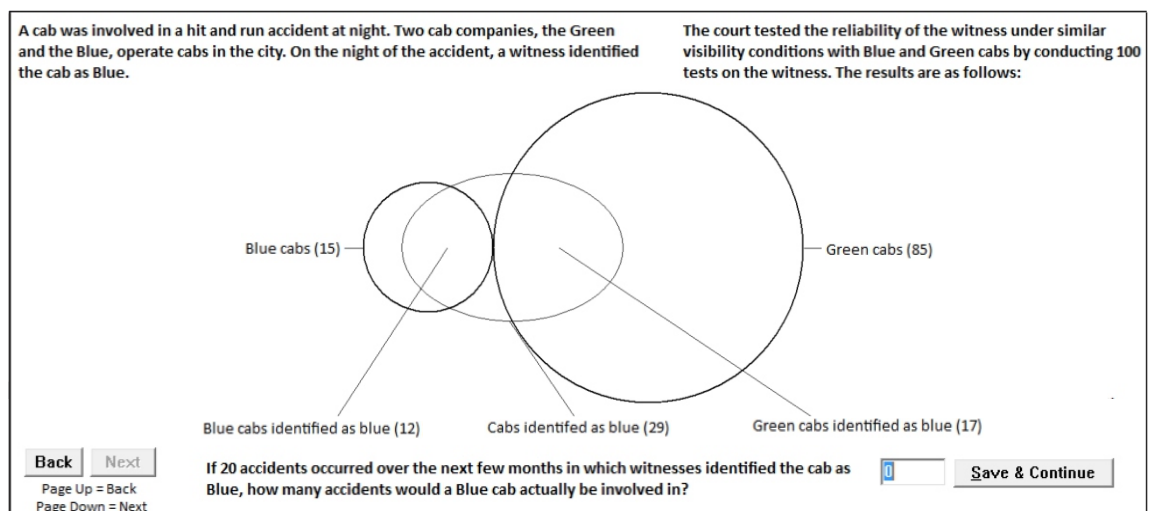
Figure 3.37: *Kellen's [2012] representation of a Bayesian problem with one question.* For the classic cab problem [Tversky and Kahneman, 1982] using natural frequencies. *Source:* [Kellen, 2012]–Figure 1
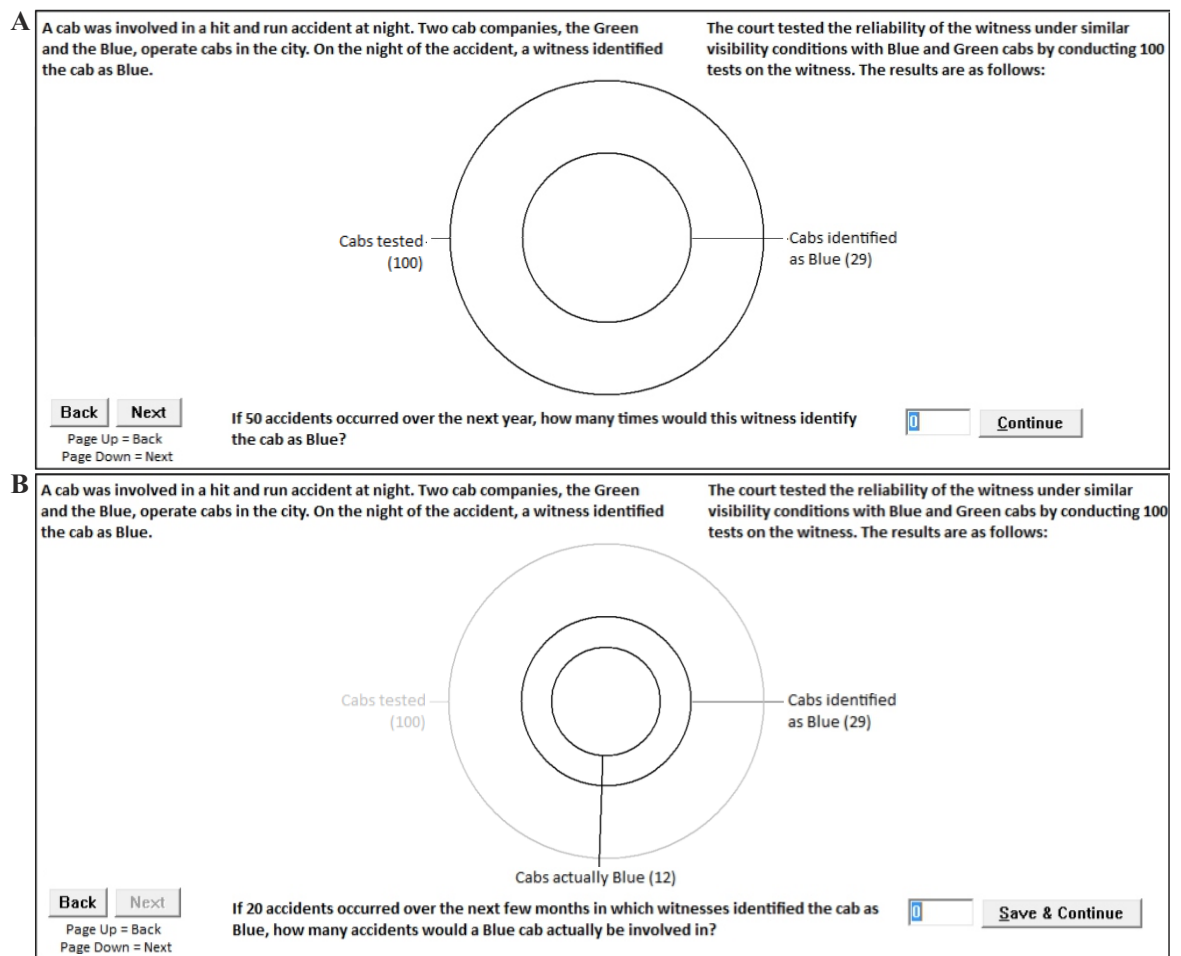
Figure 3.38: *Kellen's [2012] representation of a Bayesian problem with two questions.* Representing the classic cab problem [Tversky and Kahneman, 1982] using natural frequencies with changes to the classic text, a different type of final question (in B) and an intermediary question (in A). (A) The first visual display provided to the user. (B) The second visual display provided to the user which focuses on the sets in relation to the final question to be answered. *Source:* [Kellen, 2012]–Figure 2 and Figure 3

In contrast to previous studies, Kellen's 158 participants were diverse (university students and staff from different faculties; 57 were aged over 22; 28 did not attend high school or similar) and various Bayesian problems were tested (10 using natural frequencies: the cab [Tversky and Kahneman, 1982] and mammography [Eddy, 1982] classic problems; eight others with a base rate, hit rate and false alarm rate ranging "1-90%, 5-93% and 33-99%" and overall natural frequency population ranging from less than 50 to thousands). Thus, Kellen's findings could be more generalizable than those of other studies.

However, the final question in his problems is different from the classic one, as subjects were asked to think about future incidents rather than the present one (e.g., Kellen asked, "If 20 accidents occurred over the next few months in which witnesses identified the cab as Blue, how many accidents would a Blue cab actually be involved in?"—Figure 3.37, Figure 3.38B—but the classic problem asks "What are the chances that the cab involved in the hit-and-run accident was Blue?" [Mellers and McGraw, 1999, p. 424]). The text of his simple form (Figure 3.38) was also different as two questions had to be answered. In real-life the answer to one question is often required based on data provided statically rather than interactively. Interestingly, the quantitative data in Kellen's representations (Figure 3.37, Figure 3.38) was provided only in the diagram and so, the subjects had to parse the diagram to complete the task.

Euler diagrams visualize the nested-set relations of Bayesian problems and provide 'free-rides' (Section 3.1). They are "well-matched to meaning" [Gurr and Tourlas, 2000], and "representations are more helpful in solving statistical tasks the more closely they resemble the naturally occurring events they represent" [Sedlmeier, 1999, p. 65]. Thus Euler diagrams could facilitate Bayesian reasoning.

3.9.2 Euler Diagram with Glyphs

Representations with discrete objects can aid logical reasoning [Cosmides and Tooby, 1996]. So Euler diagrams with glyphs, as in Figure 3.39, could facilitate logical reasoning as well as reasoning about the nested-set relations of Bayesian problems, as such relations are easily depicted by Euler diagrams.
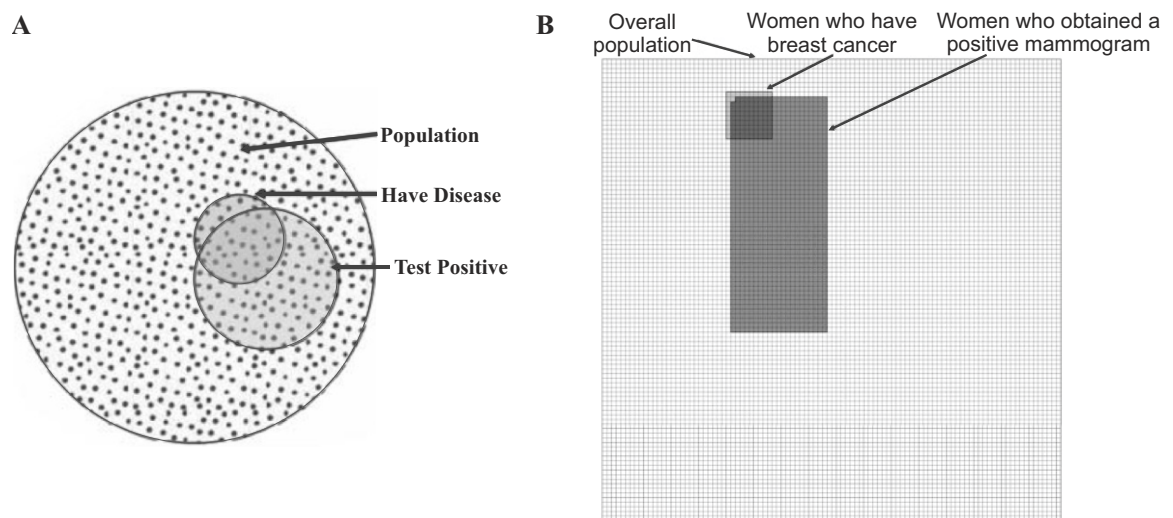
Figure 3.39: *Hybrid diagrams made up of an Euler diagram and glyphs for Bayesian reasoning.* (A) An area-proportional Euler diagram with equally-sized glyphs that should be proportional in number to the quantity assigned to the zone (for a disease diagnosis problem). *Source:* [Brase, 2009]–Figure 1 (B) An area-proportional Euler diagrams with rectilinear curves whose interior is packed with equally-sized contiguous squares that are proportional in number to the quantity assigned to the zone (for the mammography problem [Eddy, 1982]). *Source:* [Garcia-Retamero and Hoffrage, 2013]–Figure 1

Brase [2009] finds that hybrid diagrams like Figure 3.39A do not increase success compared to standard Euler diagrams or frequency grids (41.7% of 108 subjects with hybrid diagrams, 48.4% of 95 with frequency grids, 34.7% of 98 with Euler diagrams, 35.4% of 96 with text alone; non-statistically significant differences; used natural frequencies). When subjects were asked to draw their visualization as in Cosmides and Tooby's [1996] study, hybrid diagrams were even less effective (success of 28% of 50 subjects with hybrid diagrams, 49% of 49 with frequency grids, 30% of 50 with Euler diagrams).

However, Brase's [2009] tested visualization designs had a number of issues: (i) the hybrid diagram was area-proportional whereas the Euler diagram without glyphs was not, introducing a possibility of experimental confound; (ii) the number of glyphs in the hybrid diagram was inconsistent with the data (when comparing the hybrid with the standard frequency grid); (iii) the hybrid and frequency diagrams used different glyph shapes (dots versus anthropomorphic) introducing another possibility of confound.

Contrary to Brase's study, Garcia-Retamero and Hoffrage [2013] suggest that hybrid diagrams as in Figure 3.39B can help doctors and patients be more accurate when solving Bayesian problems related to

the diagnosis of different diseases (success of 62% of 40 subjects with hybrid diagrams, 26% of 40 with text alone, using natural frequencies). Doctors were more accurate than their patients (53% versus 35%), but differences were not evident when the effect of variations in the subjects' numerical abilities (i.e., the ability to handle numerical data) was controlled for. Also, patients claimed that the numerical data was not so useful when a diagram was not provided and both doctors and patients claimed that the problem was harder to complete without a diagram. Euler diagrams with curves that are not smooth, as in Garcia-Retamero and Hoffrage's, impede diagram understanding [Benoy and Rodgers, 2007] (Section 3.5.2). The curves in Brase's diagrams were regular, smooth circles, but the success of Brase's hybrid diagrams (41.7% of 108 subjects) was less than that of Garcia-Retamero and Hoffrage (62% of 40 subjecs). However, a number of differences and limitations can be noted in the two studies.

Brase's [2009] study involved a population of university psychology students with 18.7 average age who participated to partially fulfil a class requirement. Garcia-Retamero and Hoffrage's [2013] study involved 81 doctors (each paid €40) and 81 patients (each paid €20) from four hospitals in Jaén and Granada in Spain with age 18 to 85 and an even gender distribution. The subjects in both studies had an incentive and had no time constraints, which is not the case in real-life when timely, accurate decisions have to be made. Also, the populations of these studies are from the psychology and medical fields, so their findings are not fully applicable to a more diverse population of laypeople of various backgrounds.

Brase's study was conducted over only one Bayesian problem on the diagnosis of a "Disease X", which could have been hard to relate to. Garcia-Retamero and Hoffrage's study had three Bayesian problems: the mammography problem [Eddy, 1982] and two others on the diagnosis of colon cancer and insulin-dependent diabetes. Their subjects were doctors and patients with such diseases, so they could easily relate to the problems. Their findings are more generalizable than those of Brase as they had various problems, but all were about medicine with similar quantitative data. So the findings of these studies are less likely generalizable to any Bayesian problem in any application area.

Thus more studies are required to assess the effectiveness of visualizations for Bayesian reasoning.

3.10 Summary

Venn and Euler diagrams are used extensively in various application areas to visualize relationships amongst data sets. A number of automatic drawing techniques have been proposed. Most of these methods generate diagrams with the correct zones, but the curves representing the data sets are often non-smooth, not easily visible and distinguishable from other curves. Very few automatic layout techniques have been proposed to convert such diagrams into ones that are perceptually easier to comprehend, but these current techniques are computational expensive, complex and impractical for diagrams with a large number of curves. Force-directed techniques (Section 2.4.1) have been successful in laying out graphs in relatively fast time, but were never adopted to lay out Euler diagrams. In the next chapter (Chapter 4), we discuss our force-directed method, eulerForce, to improve the layout of Euler diagrams in relatively fast time.

In various application areas such as medicine and biosciences, 3-Venn diagrams that are area-proportional are often used, so that the area of the each zone in the diagram corresponds to a quantitative value in the data sets. Current automatic generation methods for such diagrams represent each data set either as a circle or as a polygon. Circles are often preferred and used in literature due to

their smoothness and continuity, but their limited degrees of freedom (i.e., a radius and a centre) impede them from accurately representing most sets of zone areas. Polygons are more flexible and are thus able to represent most sets of zone areas accurately, but their non-smooth curves lack good continuation and the curves are not easily distinguishable. Ellipses, on the other hand, are smooth like circles, but have more degrees of freedom (i.e., a centre, a semi-minor and a semi-major axis, and an angle of rotation) than circles. This motivates the need to develop drawing methods that use ellipses and to evaluate their effectiveness in drawing accurate diagrams for any set of zone areas. However, due to concerns of complexities in calculating the zone areas of overlapping ellipses and in adjusting the properties of the ellipses, none of the current drawing techniques use ellipses. In Chapter 5, we discuss our automatic drawing method, eulerAPE, to draw area-proportional 3-Venn diagrams using ellipses, together with our computational and qualitative evaluation and case study.

Though area-proportional Venn diagrams are widely used, studies suggest that humans are biased in area judgement. This is particularly problematic in areas such as Bayesian reasoning (Section 2.5) which is additionally affected by cognitive judgemental errors and biases. Thus, since studies in Bayesian reasoning illustrate that frequency grids facilitate logical reasoning (Section 2.5.5) and Euler diagrams facilitate reasoning of the data set relationships, then combining the two into a hybrid diagram, such that grids of equally-sized glyphs represent quantitative data and the curves in the Euler diagram represent the sets, might be beneficial. A few studies in psychology have been carried out to evaluate the effect of visualizations on Bayesian reasoning, but it is still unclear which is the most effective visualization. Also, most studies were carried out among a small population with a specific background (often highly educated students) and few designs, some of which were inappropriate and inconsistent with the data, were tested in each study often using just one Bayesian problem. In Chapter 6, we discuss our automatic method, eulerGlyphs, that draws frequency grids, Euler diagrams and combinations of both for Bayesian problems, together with a study which we conducted on a large, diverse population through crowdsourcing, to test a classic textual representation and six different visualizations accurately generated by eulerGlyphs for three classic Bayesian problems, and two novel textual representations which we proposed for the problem.

We will now start by discussing our force-directed layout method, eulerForce, for Euler diagrams.

Chapter 4

Force-Directed Layout for Euler Diagrams

This chapter introduces our drawing algorithm eulerForce *that lays out Euler diagrams using a force-directed approach. Its software implementation is at* `http://www.eulerdiagrams.org/eulerForce`.

4.1 Introduction

An automatic drawing algorithm should ideally draw an Euler diagram with precisely the set of zones in the abstract description, but often the generated diagrams are difficult to comprehend (Section 3.6.1) as the curves are irregular, non-smooth and not easily identified (Section 3.5.2). Algorithms that restrict the shape of the curves (e.g., to a circle) are also available (Section 3.6.2), but the generated diagrams could have empty zones (Section 3.5.2). Thus in such cases, it would be appropriate to generate a diagram that has all and only the zones defined in the abstract description and then use a layout method to transform the diagram's layout into one that satisfies specific aesthetic features. Yet only two layout methods have been proposed (Section 3.6.3) and both are complex and computationally expensive. In contrast, force-directed techniques (Section 2.4.1) have been successfully used and evaluated in areas like graph drawing to produce layouts with desired aesthetic features in relatively fast time.

In this chapter, we present our force-directed layout method, eulerForce, that produces improved layouts of Euler diagrams and in relatively fast time. Our results indicate a great potential in force-directed techniques for improving the layout of Euler diagrams and possibly for generating diagrams.

We start by outlining our motivations and objectives of this work (Section 4.2), followed by our layout method eulerForce, including the force model and the algorithm that equilibrates the forces (Section 4.3). We discuss our experimental results and evaluation of the method (Section 4.4). We end with ideas for future work (Section 4.5) and a summary of this chapter (Section 4.6).

This work has been published as a journal article [Micallef and Rodgers, 2014b] and extended abstract and poster [Micallef and Rodgers, 2009], and presented at the 2014 International Workshop on Visual Languages and Computing (VLC) and the 2009 IEEE Information Visualization (InfoVis) Poster Session. The software implementation of eulerForce is at `http://www.eulerdiagrams.org/eulerForce`.

4.2 Motivation and Objectives

Identifying the zones in the diagrams in Figure 4.1 is difficult. The diagrams were generated using a drawing algorithm [Rodgers et al., 2008b] that produces diagrams with all and only the required set of zones in the abstract description. Thus, none of the required zones are missing and none of the zones in the diagram are empty. However, the diagrams are difficult to comprehend and the zones are not easily identified, as the curves are irregular, non-smooth and not easily distinguishable (Section 3.5.2). An Euler diagram layout method could improve the diagrams' layout, but current methods (Section 3.6.3) have a number of limitations, which could be mitigated using a force-directed approach (Section 2.4.1).
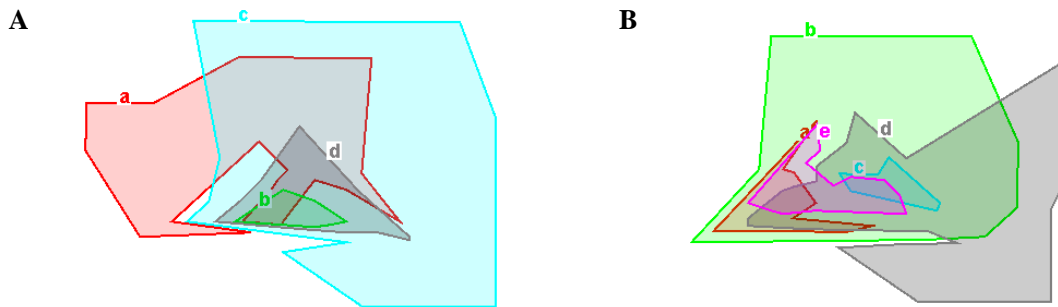
Figure 4.1: *Examples of Euler diagrams by a drawing algorithm [Rodgers et al., 2008b] that generates diagrams with all and only the zones in the given abstract description.* The diagram generated for abstract description: (A) {∅, *a*, *c*, *ac*, *cd*, *acd*, *bcd*, *abcd*}, (B) {∅, *b*, *d*, *ab*, *bd*, *be*, *abd*, *abe*, *bcd*, *bde*, *abde*, *bcde*}.

4.2.1   Limitations of Previous Euler Diagram Layout Methods

Using Graph Transformations

Rodgers et al. [2010b] defined (not implemented) a method that uses graph transformations to generate a layout that satisfies a particular wellformedness property. However, it does not take into account important curve aesthetics like regularity, smoothness and convexity and thus, it cannot improve layouts as in Figure 4.1. Also, graph transformations could be computationally expensive [Ehrig et al., 1999].

Using Multi-Criteria Optimization

Flower et al. [2003b] defined metrics for curve aesthetics (roundness, smoothness, size uniformity, closeness) to direct their optimization to an improved layout. Thus, this method could improve layouts like Figure 4.1A, but not necessarily Figure 4.1B as it handles diagrams with up to four curves. Yet, the effectiveness and correctness of these metrics were not evaluated, and assigning appropriate weights to these criteria in a multi-criteria optimization is difficult [Marler and Arora, 2004]. Helpful aesthetics like the use of regular, similarly shaped curves (Section 3.5.2) and diagram symmetry are not taken into account, and the algorithm still needs to convert the line segments in the diagram to Bézier curves after optimization to ensure curve smoothness. Also, this method uses the hill-climbing heuristic, thus it is likely to encounter a local minimum and find a local rather than global best optimized solution. More sophisticated multi-criteria optimization algorithms (e.g., simulated annealing, evolutionary algorithms) could be used, but these are computationally more expensive. In itself, multi-criteria optimization is computationally more expensive than single-criteria ones [Marler and Arora, 2004].

4.2.2 A Force-Directed Approach

Using a physical analogy of electrically charged particles (the vertices) and springs (the edges), a simple spring embedder (Section 2.4.1) converts the graph layout with multiple edge crossings in Figure 4.2A to the symmetric layout with no edge crossings, uniform edge lengths and evenly distributed vertices in Figure 4.2B. Also, this conversion is carried out in relatively fast time.
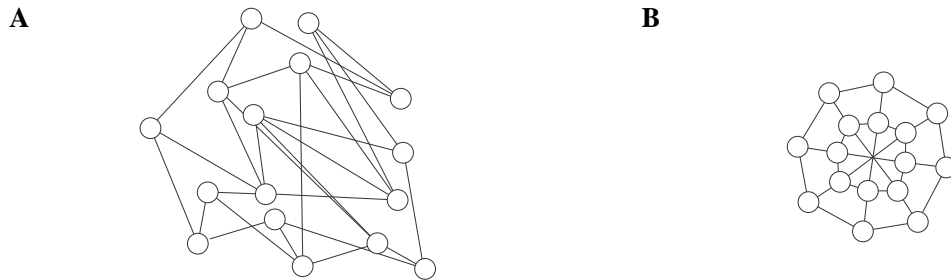
Figure 4.2: *A simple force-directed spring embedder converts the graph layout in A to that in B.* B was generated using Eades's [1984] spring embedder (which we implemented) for the graph layout in A.

A closed curve represented as a polygon is like a graph with a set of vertices and edges (Figure 4.3). So using a force-directed approach, a closed curve as in Figure 4.3A would be laid out as a smooth regular circle as in Figure 4.3B. The curve in Figure 4.3A is the same as curve *a* in Figure 4.1A. Thus, if all the curves in Figure 4.1A were converted to smooth regular circles, an easy to comprehend layout like Figure 4.4A would be produced. Similarly, Figure 4.1B could be converted to Figure 4.4B. Both diagrams in Figure 4.4 were generated using our novel force-directed layout method, eulerForce, which we discuss in Section 4.3. So a force-directed approach could be used to improve Euler diagram layouts. Also because these force-directed algorithms are often efficient, different forces could be defined to handle different constraints, and simple versions are easy to implement and to understand.
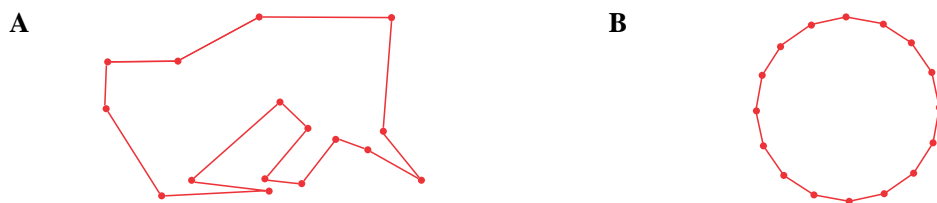
Figure 4.3: *A simple force-directed spring embedder converts the closed curve in A to that in B.* B was generated using Eades's [1984] spring embedder (which we implemented) for the graph layout in A.
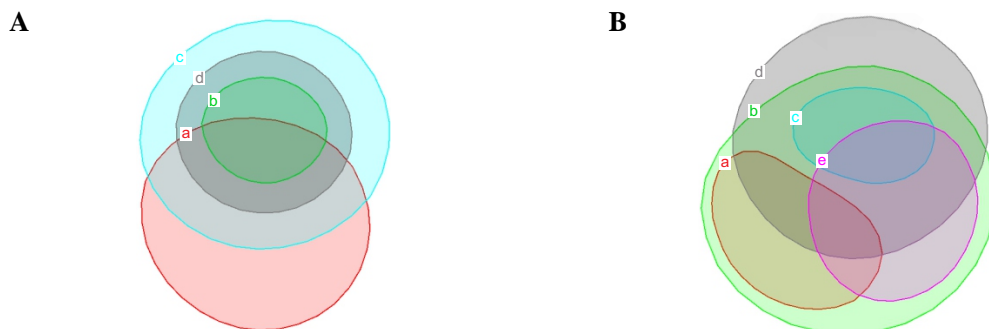
Figure 4.4: *The improved layouts generated by our force-directed spring embedder,* eulerForce, *for the Euler diagrams in Figure 4.1.* The improved layout for: (A) Figure 4.1A, (B) Figure 4.1B.

To successfully adopt a force-directed approach to improve the layouts of Euler diagram, techniques studied and used in graph drawing (Section 2.4.1) have to be investigated. A major challenge that is not a problem in graph drawing is the development of an appropriate force model that, besides satisfying the required aesthetic criteria, maintains the zones in the original layout, in that no zones are added or lost during the layout improvement process. Thus, initially, simple force-directed methods such as the spring embedder should be investigated. This method is easy to understand and to implement and so, the required forces for Euler diagram layouts can be studied including an investigation of how these forces interact with one another. Once this is understood and an appropriate simple force model is defined, more complex force-directed techniques that improve performance and handle more aesthetic criteria and complex large Euler diagrams with various vertices and edges can be investigated.

4.2.3 Our Objectives

We devised the layout method *eulerForce* to improve the layout of automatically drawn Euler diagrams that depict all and only the zones in the abstract description, like those generated by Rodgers et al.'s [2008b] method. This is the first force-directed layout method for Euler diagrams, so **we aimed to improve important diagram aesthetics** that could otherwise impede understanding (Section 3.5). Yet we still wanted to **depict the same set of zones as in the initial layout** (the layout before the layout improvement process commences). Thus, we devised a force model that:

1. Obtains regular, smooth and similarly shaped convex curves;
2. Maintains the same set of zones as that in the initial diagram layout;
3. Ensures zones are connected (i.e., made up of one region);
4. Ensures curves are not too close to one another;
5. Ensures contained curves are centred in their containing curve or zone;
6. Attains adequately sized zones and curves.

By satisfying these objectives, the final layout is easy to comprehend, and the curves and the zones are distinguishable and easily identified (Section 3.5). Example, Benoy and Rodgers 's [2007] study shows that diagrams that do not satisfy the curve aesthetics in **objectives 1, 4 and 6**, the diagram is difficult to understand. The force model also aims to obtain layouts that are wellformed (Section 3.4.4) as this would also aid diagram comprehension (Section 3.5.1). In particular, **objective 3 and 4** ensure that the most important wellformedness properties, that of connected zones and discrete curves, are met—Rodgers et al.'s [2012b] study indicates that diagrams that do not satisfy any one of these two properties are the most difficult to use. Final layouts are also highly likely to be symmetric [Eades and Lin, 2000], particularly when **objective 5** is met (as Figure 4.4A). Thus using such layout algorithms, automatically generated diagrams depicting all and only the required zones (as Figure 4.1) could be used instead of diagrams with empty zones generated by algorithms that restrict the shape of the curve.

We aspired to develop a simple yet appropriate force model for Euler diagrams that could later be extended to handle other criteria. Thus we aimed at using a simple force-directed algorithm that aids understanding of the forces and how the different forces interact with one another. Various techniques could be adopted to improve the performance of such simple methods, overcome local minima and handle more criteria and more complex large systems with various vertices and edges (Section 2.4.1). Based on our findings, we wanted to identify a number of techniques that should be considered next to ensure the fast and efficient generation of comprehensible Euler diagram layouts for any number of curves and set of zones. We discuss our force model and algorithm in the next section (Section 4.3).

4.3 The Layout Method – eu*lerForce*

Each curve in the given Euler diagram is represented as a ***polygon*** with a number of vertices and edges. A force model acts on the vertices, edges and entire polygons to improve the diagram's layout while still depicting the same set relations, so the objectives in Section 4.2.3 are reached. Thus, given a diagram, the algorithm iteratively applies forces in our model on the diagram's layout based on its current state. The final improved layout is returned once the forces equilibrate and no further changes are made.

4.3.1 The Force Model

Our physical system is similar to that of a simple spring embedder (Section 2.4.1) as shown in Figure 4.5.
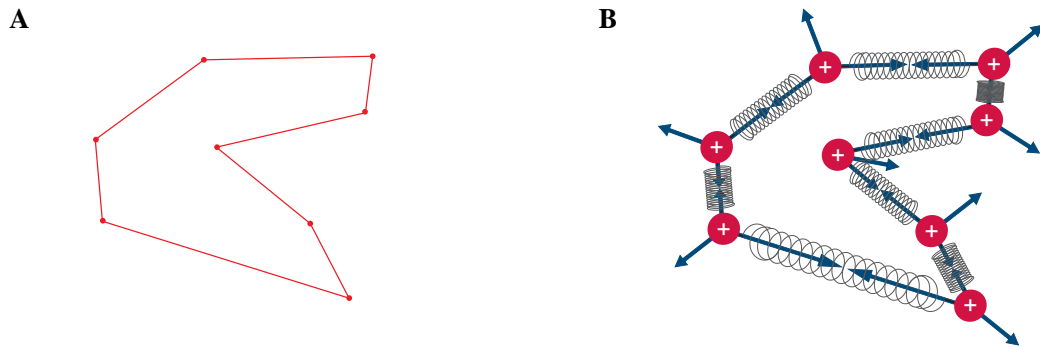
Figure 4.5: *The physical system and forces in a simple spring embedder.* (A) Vertices and edges defining a polygon, possibly a curve in an Euler diagram. (B) The vertices in A act like electrically charged particles that repel one another. The edges in A act like springs that attract particles attached to their endpoints.

Two Types of Forces

Our force model consists of ***repulsive*** and ***attractive*** forces between different ***structures*** in the layout, including: (i) vertices, (ii) edges, (iii) entire polygons. Thus, our forces differ from those used in simple graph drawing methods by systematically moving any of these structures and not only the vertices.

Similar to the typical spring embedder in graph drawing, our repulsive forces follow the inverse square law and our attractive forces follow the Hooke's law [Di Battista et al., 1999a]. So, given ***d* is the Euclidean distance between two structures** s_1 **and** s_2 **in the layout**, these forces are defined as:

> ***Repulsive Forces***
>
> inversely proportional to the squared distance between structures s_1 and s_2,
> so the repulsive force between s_1 and s_2, that is,
> the repulsive force exerted on s_2 by s_1 and on s_1 by s_2 is $f_r = c_r/d^2$
> where c_r is a constant that determines the strength of the force;

> ***Attractive Forces***
>
> directly proportional to the distance between structures s_1 and s_2,
> so the attractive force exerted between s_1 and s_2, that is,
> the attractive force exerted on s_1 and s_2 by the spring between s_1 and s_2 is $f_a = c_a d$
> where c_a is the stiffness of the spring that determines the strength of the force and the natural length of the spring is zero.

Each repulsive and attractive force in our model is specifically devised to reach one of the objectives in Section 4.2.3. So as shown in the next sections, the involved structures and the value of constants c_r and c_a are not the same for all the forces. The **value of the predefined (hardcoded) c_r or c_a for each force** was determined after the layout improvement process was observed for a number of different diagrams using different values for c_r and c_a of our forces. The chosen values were the ones that best met our objectives for the tested diagrams. The **direction of a force** in our model is **always** as defined above for of its type. The **magnitude of a force** is also as defined above for its type, **except for specific cases** were the magnitude (never the direction) is defined differently. Such special repulsive and attractive forces are respectively indicated in the next sections using ◄●●►* and ●►◄●*.

Our repulsive forces are the same as those used in Eades's [1984] spring embedder. Our attractive forces are different from those of Eades, as Eades uses logarithmic rather than linear (Hooke's law) springs stating that the latter could be too strong. However, Di Battista et al. argue that, "it is difficult to justify the extra computational effort by the quality of the resulting drawings" [Di Battista et al., 1999a, p. 309]. Since our attractive forces assume linear, Hooke's law springs with natural length zero, they are the same as those used in Tutte's [1963] barycentre method. We opted for such attractive forces, as these forces are used to smooth the curves and to regain zones that are lost during the layout improvement process. Thus, while in the former the edges should be as short as possible to produce smooth curves, in the latter the force of the spring should be strong enough to attract structures and regain the lost zones.

We will now discuss how such repulsive and attractive forces between vertices, edges and polygons are used in our force model to generate layouts that meet our objectives in Section 4.2.3.

Forces to Meet Objective 1

To ***obtain regular, smooth and similarly shaped convex curves*** (objective 1, Section 4.2.3), we use the following forces that are typical for a simple spring embedder [Di Battista et al., 1999a]:

> (**F1**) **Repulsion for vertices not to be too close to one another**   ◄●●►
>
>> For every polygon p in the current layout and
>> for every pair of distinct vertices v_1 and v_2 of p,
>> a *repulsive force* with $c_r = 3500$ is exerted between v_1 and v_2,
>> so that v_1 and v_2 move away from one another;
>
> (**F2**) **Attraction for approximately uniform edge lengths**   ●►◄●
>
>> For every polygon p in the current layout and
>> for every pair of distinct vertices v_1 and v_2 of p that are connected by an edge,
>> an *attractive force* with $c_a = 0.5$ is exerted between v_1 and v_2,
>> so that v_1 and v_2 move closer to one another.

Starting with a non-smooth polygon (Figure 4.6A), the vertices (acting as electrically charged particles) of the polygon repel one another and the edges (acting as springs) attract the vertices connected to their endpoints (Figure 4.6B). So when the forces equilibrate, a regular, smooth and convex polygon is obtained (Figure 4.6C). In this way, these forces will also help the algorithm achieve similarly shaped curves, as in for instance Figure 4.4A.
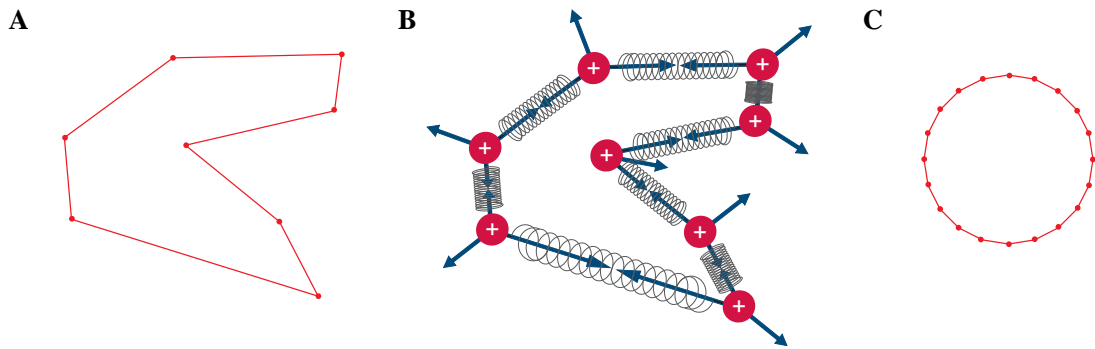
Figure 4.6: *Forces to obtain regular, smooth and similarly shaped convex curves (objective 1).* (A) The polygon representing a curve in a given diagram. (B) The physical system with forces F1 and F2 between the vertices of the polygon in A. (C) The final polygon layout once the forces in B equilibrate.

Forces to Meet Objective 2

To **maintain the same set of zones as that in the initial diagram layout** (objective 2, Section 4.2.3), so that no zones are added or lost during the layout improvement process, we devised a set of forces for each different type of curve relation to ensure that: (a) the current improved layout maintains the zones in the initial layout; (b) if the current layout has new zones or is missing any of the zones in the initial layout, forces correct the layout accordingly. We opted to use forces to correct layouts that depict the incorrect set of zones rather than to disallow such layouts altogether, to avoid local minima. So for every pair of distinct polygons in the initial layout, the following forces are applied:

> (**F3**) If the two polygons in the **initial layout** *do not intersect*
> and in the **current layout** they still *do not intersect* (Figure 4.7A):
>
>> If p_1 and p_2 are these two polygons in the current layout,
>> for every vertex v_1 of p_1 and for every vertex v_2 of p_2
>> (e.g., Figure 4.7B),
>>
>>> a *repulsive force* with $c_r = 1500$ is exerted between v_1 and v_2,
>>
>> so that these vertices move accordingly and
>> the required disjointness of p_1 and p_2 is reinforced (Figure 4.7C);
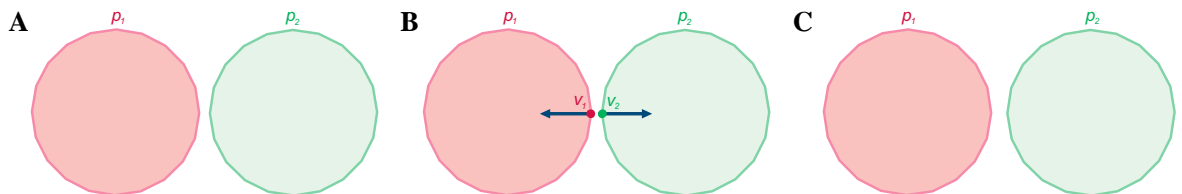
Figure 4.7: *Forces to maintain the same zones as those in the initial layout (objective 2)—if, consistent with the initial layout, two polygons in the current layout do not intersect, the vertices of the two polygons repulse to reinforce the required disjointness.* (A) The two polygons, p_1 and p_2, in the current layout. (B) The repulsive force between a vertex v_1 of p_1 and a vertex v_2 of p_2. (C) The polygons once the repulsive force is exerted between the vertices of p_1 and p_2 and these vertices are moved accordingly.

(**F4**) If the two polygons in the **initial layout** *do not intersect*

but in the **current layout** they *do intersect* (e.g., Figure 4.8A):

If p_1 and p_2 are these two polygons in the current layout,

for every vertex v_1 of p_1 and vertex v_2 of p_2

if v_1 is inside or on an edge of p_2 and v_2 is inside or on an edge of p_1 (e.g., Figure 4.8B),

an *attractive force* with $c_a = 0.2$ is exerted between v_1 and v_2,

if v_2 is not inside or on an edge of p_1 (e.g., Figure 4.8C-D),

a *repulsive force* with $c_r = 1500$ is exerted on v_1 by v_2,

if v_1 is not inside or on an edge of p_2 (e.g., Figure 4.8E-F),

a *repulsive force* with $c_r = 1500$ is exerted on v_2 by v_1,

so that these vertices move accordingly and

the required disjointness of p_1 and p_2 is regained (as in Figure 4.7A);
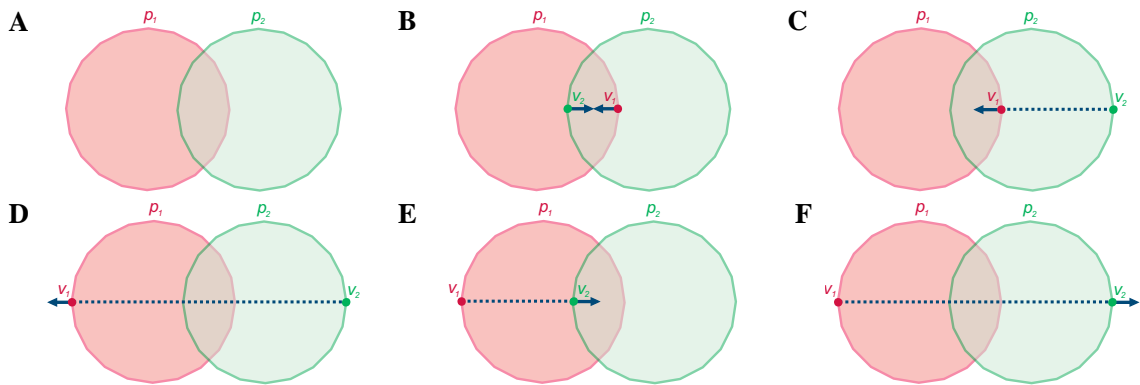
Figure 4.8: *Forces to maintain the same zones as those in the initial layout (objective 2)—if, inconsistent with the initial layout, two polygons in the current layout intersect, attractive and repulsive forces between specific vertices of the two polygons strive to regain the required disjointness.* (A) An example of how the two polygons, p_1 and p_2, in the current layout could be inconsistent with the initial layout in that they intersect. (B) The attractive force between a vertex v_1 of p_1 and a vertex v_2 of p_2, when v_1 is inside or on an edge of p_2 and v_2 is inside or on an edge of p_1. (C) The repulsive force on v_1 by v_2, when v_2 is not inside or on an edge of p_1—in this case, v_1 is inside p_2. (D) The same type of force as in C, but here v_1 is not inside or on an edge of p_2. (E) The repulsive force on v_2 by v_1, when v_1 is not inside or on an edge of p_2—in this case, v_2 is inside p_1. (F) The same type of force as in E, but here v_2 is not inside or on an edge of p_1.

(**F5**) If the two polygons in the **initial layout** *intersect*

and in the **current layout** they still *intersect* (Figure 4.9A):

If p_1 and p_2 are these two polygons in the current layout,

for every vertex v_1 of p_1 and for every vertex v_2 of p_2,

if both v_1 and v_2 are on the boundary of the overlapping region,

that is v_1 is inside p_2 and v_2 is inside p_1, (e.g., Figure 4.9B),

a *repulsive force* with $c_r = 1000$ is exerted between v_2 and v_1,

so that these vertices move accordingly and

the required intersection of p_1 and p_2 is reinforced (Figure 4.9E);

if v_1 is not inside p_2 and v_2 is inside or on an edge of p_1
(e.g., Figure 4.9C),

a *repulsive force* with $c_r = 1000$ is exerted on v_1 by v_2,

so that these vertices move accordingly and

p_1 and p_2 are not too close to one another (Figure 4.9E);

if v_2 is not inside p_1 and v_1 is inside or on an edge of p_2
(e.g., Figure 4.9D),

a *repulsive force* with $c_r = 50$ is exerted on v_2 by v_1,

so that these vertices move accordingly and

p_1 and p_2 are not too close to one another (Figure 4.9E);
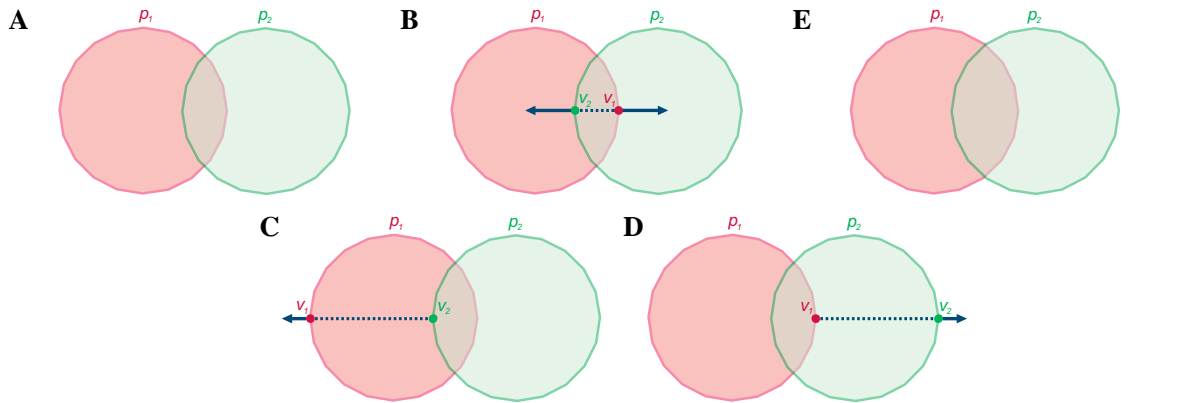
Figure 4.9: *Forces to maintain the same zones as those in the initial layout (objective 2)—if, consistent with the initial layout, two polygons in the current layout intersect, the vertices of the two polygons that are on the boundary of the overlapping region repulse to reinforce the required intersection.* (A) The two polygons, p_1 and p_2, in the current layout. (B) The repulsive force between a vertex v_2 of p_2 and a vertex v_1 of p_1, where both v_1 and v_2 are the boundary of the overlapping region. (C) The repulsive force on a v_1 by a v_2, when v_1 is not in p_2 but v_2 is in or on an edge of p_1. (D) The repulsive force on a v_2 by a v_1, when v_2 is not in p_1 but v_1 is in or on an edge of p_2. (E) The polygons once the repulsive force is exerted between the vertices of p_1 and p_2.

(**F6**) If the two polygons in the **initial layout** *intersect*

but in the **current layout** they *do not intersect* (e.g., Figure 4.10A):

If p_1 and p_2 are these two polygons in the current layout,
for every vertex v_1 of p_1 and vertex v_2 of p_2
(e.g., Figure 4.10B-C)

a *special attractive force* defined as $f = c/d^2$

where c is a constant (0.1) determining the strength of the force
and d is the Euclidean distance between v_1 and v_2

is exerted between v_1 and v_2,

so that these vertices move accordingly and
the required intersection of p_1 and p_2 is regained (as in Figure 4.9A);
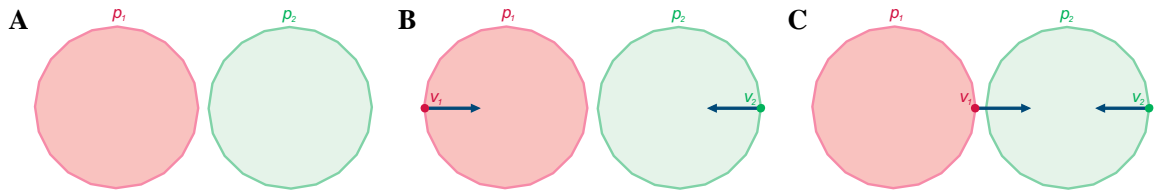
Figure 4.10: *Forces to maintain the same zones as those in the initial layout (objective 2)—if, inconsistent with the initial layout, two polygons in the current layout do not intersect, a special attractive force that is inversely proportional to the squared distance between vertices of the two polygons strives to regain the required intersection.* (A) Example of how two polygons, p_1 and p_2, in the current layout could be inconsistent with the initial layout as they do not intersect. (B) The special attractive force between a vertex v_1 of p_1 and a vertex v_2 of p_2. (C) Same force as in B but between different vertices.

(**F7**) If, in the **initial layout**, one of the polygons *contains* the other
and in the **current layout**, the polygons still *depict* **the required** *containment*:

> If p_1 and p_2 are these two polygons in the current layout
> and p_2 is contained in p_1 (Figure 4.11A),
> for every vertex v_1 of p_1 and for every vertex v_2 of p_2
> (e.g., Figure 4.11B),
>
> > a *repulsive force* with $c_r = 1000$ is exerted between v_1 and v_2,
>
> so that these vertices move accordingly and
> the required containment of p_2 in p_1 is reinforced (Figure 4.11C);
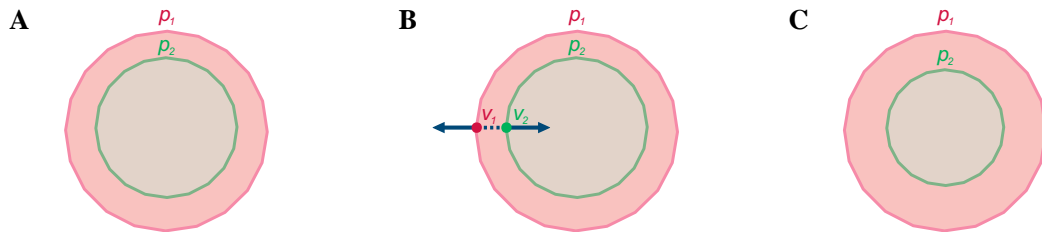
Figure 4.11: *Forces to maintain the same zones as those in the initial layout (objective 2)—if, consistent with the initial layout, two polygons in the current layout depict the required containment, the vertices of the two polygons repulse to reinforce the required containment.* (A) The two polygons, p_1 and p_2, in the current layout, where p_1 contains p_2. (B) The repulsive force between a vertex v_1 of p_1 and a vertex v_2 of p_2. (C) The polygons once the repulsive force is exerted between the vertices of p_1 and p_2.

(**F8**) If, in the **initial layout**, one of the polygons *contains* the other
but in the **current layout**, the polygons *do not depict* **the required** *containment*
(e.g., Figure 4.12A):

> If p_1 and p_2 are these two polygons in the current layout
> and according to the initial layout, p_2 should be contained in p_1 (as in Figure 4.11A),
> for every vertex v_1 of p_1 and vertex v_2 of p_2
> if v_1 is inside or on an edge of p_2 and v_2 is not inside or on an edge of p_1
> (e.g., Figure 4.12B),
>
> > an *attractive force* with $c_a = 0.15$ is exerted between v_2 and v_1,

if v_2 is inside or on an edge of p_1 (e.g., Figure 4.12C-D),

    a *repulsive force* with $c_r = 2000$ is exerted on v_1 by v_2,

if v_1 is not inside or on an edge of p_2 (e.g., Figure 4.12E-F),

    an *attractive force* with $c_a = 0.1$ is exerted on v_2 from v_1,

so that these vertices move accordingly and

the required containment of p_2 in p_1 is regained (as in Figure 4.11A).
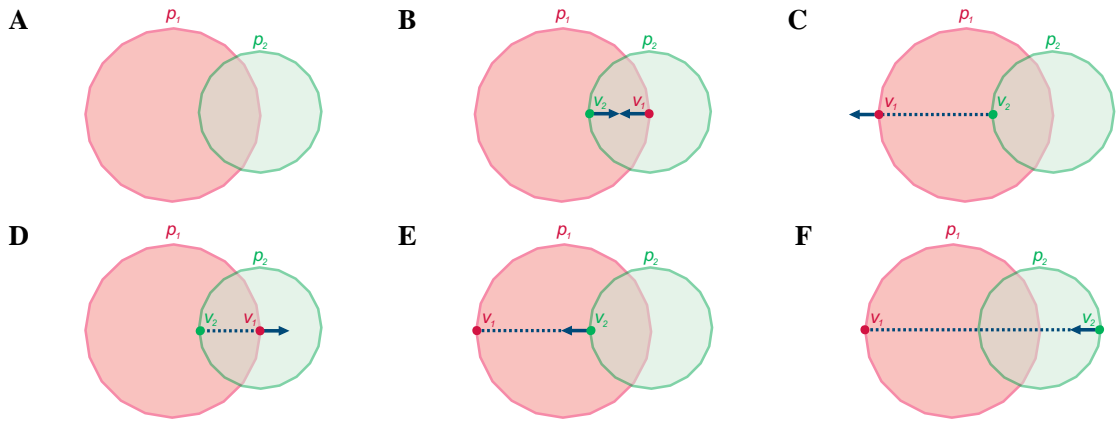
Figure 4.12: *Forces to maintain the same zones as those in the initial layout (objective 2)—if, inconsistent with the initial layout, two polygons in the current layout do not depict the required containment, attractive and repulsive forces between specific vertices of the two polygons strive to regain the required containment.* (A) An example of how the two polygons, p_1 and p_2, in the current layout could be inconsistent with the initial layout in that p_1 does not contain p_2. (B) The attractive force between a vertex v_1 of p_1 and a vertex v_2 of p_2, when v_1 is inside or on an edge of p_2 and v_2 is not inside or on an edge of p_1. (C) The repulsive force on v_1 by v_2, when v_2 is inside or on an edge of p_1 and in this case, v_1 is not inside p_2. (D) The same type of force as in C, but in this case, v_1 is inside or on an edge of p_2. (E) The repulsive force on v_2 by v_1, when v_1 is not inside or on an edge of p_2 and in this case, v_2 is inside p_1. (F) The same type of force as in E, but in this case, v_2 is not inside or on an edge of p_1.

In the above, forces F3 to F8 are applied between vertices of polygons to (a) maintain the zones of the initial layout and (b) correct layouts that are not depicting the same set of zones as that of the initial layout. However, to ensure (a) and reduce the need for (b), if a vertex v_1 of polygon p_1 is closer to a point x on an edge $e = (v_{2i}, v_{2j})$ of a polygon p_2 than vertex v_{2i} or v_{2j} of p_2 (as in Figure 4.13), **forces F3-F8 are also applied between v_1 and e**, such that e is moved based on the forces exerted on it about x.
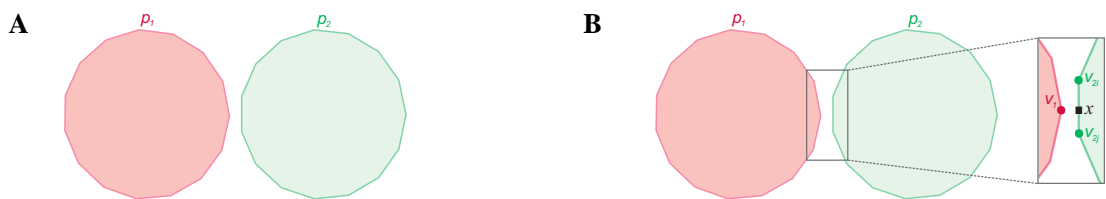
Figure 4.13: *Forces to maintain the same zones as those in the initial layout (objective 2)—forces F3 to F8 are also applied between vertices and edges when the vertex of one polygon is closer to a point on the edge of a second polygon than the vertices of this second polygon.* (A) Two polygons, p_1 and p_2, in a layout generated at some point during the layout improvement process. (B) A magnified view of a section of the layout. Vertex v_1 of p_1 is closer to point x on the edge (v_{2i}, v_{2j}) of p_2 than to vertices v_{2i} and v_{2j}.

Forces to Meet Objective 3

As the vertices are moved during the layout improvement process, zones could become disconnected making the diagram layout non-wellformed and thus difficult to comprehend (Section 3.5.1).

To ***ensure zones are connected, so all are made up of one region*** (objective 3, Section 4.2.3), the following force is applied to layouts with at least one disconnected zone, just before any of the forces that strive to meet objective 2 are considered. So, for every pair of distinct polygons, p_1 and p_2, in the **current layout** and for every zone z in any or both of p_1 and p_2:

> (**F9**) While **zone z is *disconnected*** (e.g., Figure 4.14A)
>
>> If r is the smallest region of z,
>>
>> for every vertex v_1 of p_1 and vertex v_2 of p_2
>>
>> if v_1 is inside or on an edge of r and v_2 is not inside or on an edge of r (e.g., Figure 4.14B-C),
>>
>>> an *attractive force* with $c_a = 0.05$ is exerted between v_1 and v_2,
>>
>> so that these vertices move accordingly and
>>
>> a region from z is discarded (as in Figure 4.14D).
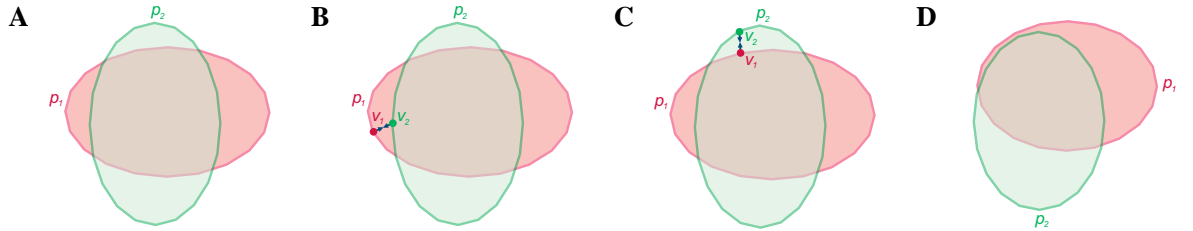
Figure 4.14: *Forces to ensure zones are connected, so all are made up of one region (objective 3).* (A) An example of two polygons, p_1 and p_2, with two disconnected zones whose regions are precisely inside one of the polygons. (B) The attractive force between a vertex v_1 of p_1 and a vertex v_2 of p_2, when v_1 and v_2 are inside or on an edge of the smallest region of a disconnected zone (here the disconnected zone has two regions in exactly p_1). (C) The same type of force as in B, but for a different disconnected zone (here the disconnected zone has two regions in exactly p_2). (D) The polygons with connected zones once the attractive force is exerted between the relevant vertices of p_1 and p_2 and these vertices are moved accordingly.

Forces to Meet Objective 4

Layouts with curves close to one another are difficult to comprehend [Benoy and Rodgers, 2007] and could break the important wellformedness property of discrete curves [Rodgers et al., 2012b]. Repulsive forces ensure that vertices (or other structures) are not too close to one another. As the vertices move by the force, the edges connected to them also move. So most of the previously discussed repulsive forces will also ***ensure curves are not too close to one another*** (objective 4, Section 4.2.3). This is particularly relevant to the forces for objective 2, as in some cases they are also applied between vertices and edges.

Forces to Meet Objective 5

Sometimes a curve is contained in another curve or zone. The repulsive forces in the model would ensure that this contained polygon remains inside the containing polygon or zone. However, centring this contained polygon in its containing polygon or zone, so that its boundary is equidistant from that of

the containing structure, could improve the layout and its symmetry. To ***ensure contained curves are centred in their containing curve or zone*** (objective 5, Section 4.2.3), the following force is applied:

> (**F10**) When a **polygon is contained in another polygon or zone** (e.g., Figure 4.15A, D)
>
>> If c_1 is the centroid of the contained polygon
>> and c_2 is the centroid of the containing polygon or zone
>> (e.g., Figure 4.15B, E),
>>
>>> an *attractive force* with $c_a = 0.05$ is exerted on c_1 from c_2,
>>
>> so that the entire contained polygon is moved closer to c_2 and
>> centred in its containing polygon or zone (e.g., Figure 4.15C, F).
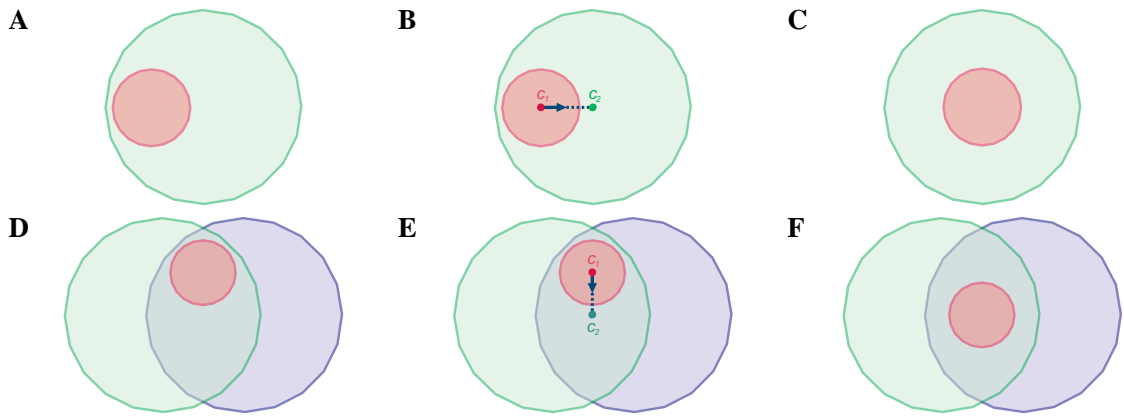
Figure 4.15: *Forces to centre contained curves in their containing curve or zone (objective 5).* (A) An example of a polygon contained in another polygon. (B) The attractive force on centroid c_1 of the contained polygon from centroid c_2 of the containing polygon. (C) The layout once the attractive force is exerted. (D) An example of a polygon contained in a zone. (E) The attractive force on centroid c_1 of the contained polygon from centroid c_2 of the containing zone. (F) The layout once the attractive force is exerted.

Forces to Meet Objective 6

If the size of the zones is inadequate, the layout could be difficult to understand, particularly when zones are not easily visible and their area is disproportional to that of other zones [Benoy and Rodgers, 2007]. An adequate zone area could be one that is similar to the area of other zones in the layout, so that the diagram area is evenly distributed among its zones [Benoy and Rodgers, 2007].

Another option is for the zone area to be inversely proportional to the number of curves in which it resides, in that the greater the number of curves it is located in, the smaller the zone area [Stapleton et al., 2012]. So if a k-curve zone is a zone that is located in k curves of a diagram with n curves, the area of the zone is assigned a weight $w = n/k$. Example, if a diagram has three curves ($n=3$), a 1-curve zone ($k=1$; $w=3$) will be twice as large as a 2-curve zone ($k=2$; $w=3/2$) and three times as large as a 3-curve zone ($k=3$; $w=1$). This ensures that outermost curves in the diagram are similar in size even when any one of the curves contains other curves. This is illustrated in Figure 4.16 where the difference in the area of curves a and b in Figure 4.16A, in which layout the area of the zones is inversely proportional to the number of curves the zones reside in, is smaller than that in Figure 4.16B where all the zones have the same area. Also, for layouts to have same sized zones, the shape of some of the curves might have to be less regular than desired and contained curves might not fit in the containing curve or zone. Example, the curve labelled c in Figure 4.16B is depicted by an ellipse rather than a circle. If curve c is depicted

as a circle with the same area as that of the ellipse in Figure 4.16B, curve c would intersect with curve b. This issue is not visible in Figure 4.16A, as the area of zone ac, thus the area of curve c, is smaller than that of zone a, as zone ac is in two curves rather than one. Flower et al.'s [2003b] layout method aims at having same sized zones. However, the authors indicate that often this is not possible as other important curve aesthetics such as curve smoothness and closeness could be compromised. Thus, in this work, an adequate area for a k-curve zone in a diagram with n curves is proportional to $w = n/k$.
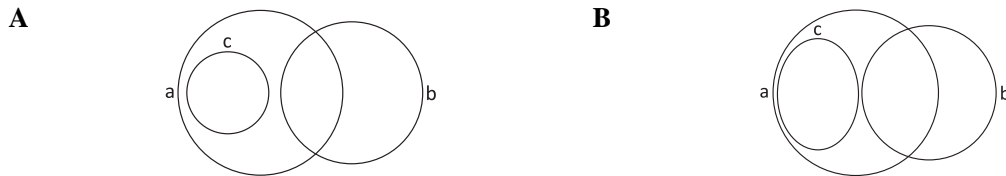
Figure 4.16: *Possible ways how to adequately size zones and curves in a diagram.* (A) The area of each zone is the inversely proportional to the number of curves the zone is located in. (B) The area of all the zones is equal. However, curve c is depicted as an ellipse, as a circle with the required area would not fit in zone a.

To ***attain the adequately sized zones and curves*** (objective 6, Section 4.2.3), a set of forces are required to change the size of the polygons depicting the curves and to move these polygons closer or further away from one another, so that the required adequate zone areas are obtained. The size of the polygons is changed appropriately by progressively changing the strength of the repulsive force F1 that ensures that vertices of polygons are not too close to one another. The greater the repulsive force, the further away neighbouring vertices of a polygon are from one another, thus enlarging the polygon. Once the size of the polygons is changed so their area is the sum of the required adequate area of the zones they contain, the polygons are moved using the following forces to get the required zone areas:

(**F11**) To **increase the area of a zone** (e.g., Figure 4.17A)

> If z is the zone whose area should be increased
> and c_1 is the centroid of z,
> for every polygon p that contains z,
> if c_2 is the centroid of p (e.g., Figure 4.17B, C),
>> an *attractive force* with $c_a = 0.03$ is exerted on c_2 from c_1,
> so that the entire polygon p is moved closer to c_1,
> thus increasing its size (e.g., Figure 4.17D).
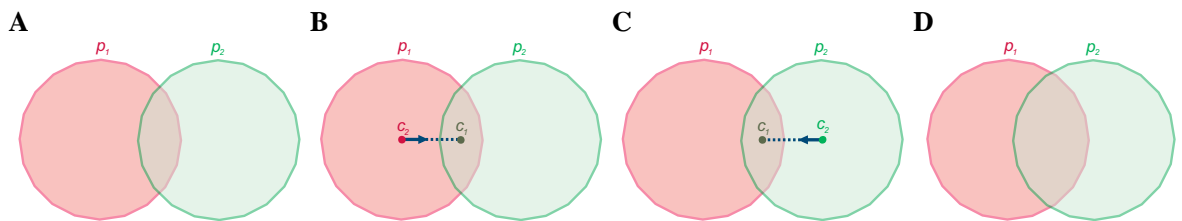
Figure 4.17: *Forces to attain adequately sized zones and curves (objective 6)—increasing the area of a zone.* (A) In this example, the area of the zone that is in both polygon p_1 and p_2 (compromised by the overlapping region) should be increased. (B) The attractive force on the centroid c_2 of p_1 from the centroid c_1 of the zone to be enlarged. (C) The attractive force on the centroid c_2 of p_2 from the centroid c_1 of the zone to be enlarged. (D) The layout once the attractive force is exerted on the centroid of all the polygons that contain the zone from the centroid of the zone and the entire polygons are moved accordingly.

(**F12**) To **decrease the area of a zone** (e.g., Figure 4.18A)

    If z is the zone whose area should be decreased

    and c_1 is the centroid of z,

    for every polygon p that contains z,

    if c_2 is the centroid of p (e.g., Figure 4.18B, C),

       a *repulsive force* with $c_r = 4000$ is exerted on c_2 from c_1,

    so that the entire polygon p is moved further away from c_1,

    thus decreasing the size of z (e.g., Figure 4.18D).
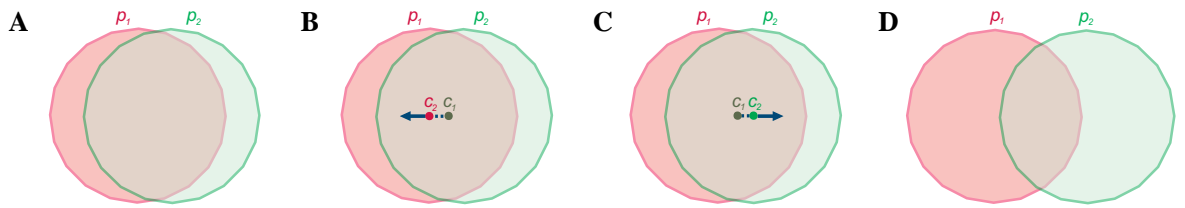
Figure 4.18: *Forces to attain adequately sized zones and curves (objective 6)—decreasing the area of a zone.* (A) Here the area of the zone that is in both polygon p_1 and p_2 should be decreased. (B) The repulsive force on centroid c_2 of p_1 from centroid c_1 of the zone to be enlarged. (C) The repulsive force on centroid c_2 of p_2 from centroid c_1 of the zone to be enlarged. (D) The layout once the repulsive force is exerted on the centroid of all polygons that contain the zone from the centroid of the zone and the entire polygons moved.

Similar to the forces for objective 2 (i.e., maintaining the same zones as in the initial layout), other forces besides F11 and F12 have been added to correct generated layouts whose zones are not the same as in the initial layout. We do not disallowed incorrect layouts to reduce the chances of reaching a local minimum. So, if while increasing or decreasing zone areas,

(**F13**) The **current layout has a zone that is not depicted in the initial layout**

    (e.g., Figure 4.19A is the initial layout, Figure 4.19B is the current layout)

    If z is the zone that is in the current but not the initial layout

    and c_1 is the centroid of z,

    for every polygon p that contains z in the current but not in the initial layout,

    if c_2 is the centroid of p (e.g., Figure 4.19C),

       a *repulsive force* with $c_r = 2000$ is exerted on c_2 from c_1,

    so that the entire polygon p is moved further away from c_1,

    thus reducing the size of z and its appearance in the layout

    until it is no longer visible (as in Figure 4.19A).
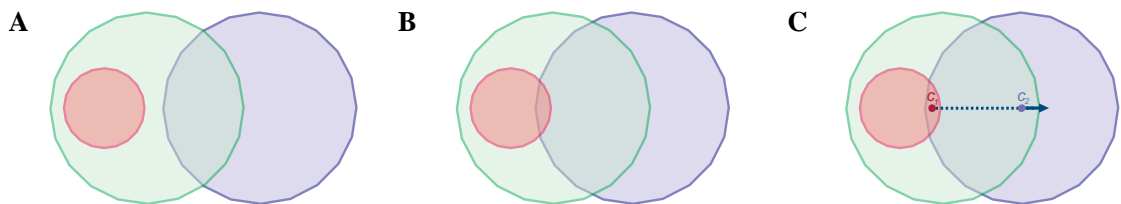
Figure 4.19: *Forces to attain adequately sized zones and curves (objective 6)—correcting layouts with a new zone.* (A) An initial layout. (B) The current layout with a zone that is not in the initial layout. (C) The repulsive force on centroid c_2 of p_1 from centroid c_1 of the unwanted zone, which moves p_1 away from c_1.

(**F14**) The **current layout does not have a zone that is depicted in the initial layout**

(e.g., Figure 4.20A is the initial layout, Figure 4.20B is the current layout)

If z is the zone that is in the initial but not the current layout,

for every pair of distinct polygons p_1 and p_2 that should contain z,

if c_1 is the centroid of p_1 and c_2 is the centroid of p_2 (e.g., Figure 4.20C),

an *attractive force* with $c_a = 0.05$ is exerted between c_1 and c_2,

so that the entire polygons p_1 and p_2 are moved closer to one another
and once all the distinct pairs of polygons that should contain z get closer,
the missing zone is regained (as in Figure 4.20A).

A   **B**   **C**
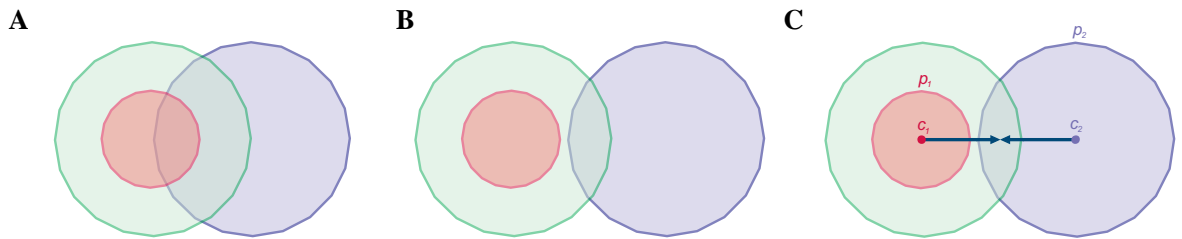
Figure 4.20: *Forces to attain adequately sized zones and curves (objective 6)—correcting layouts with a missing zone.* (A) The initial layout. (B) The current layout with a missing zone that is in the initial layout. (C) Attractive force between centroid c_1 of p_1 and centroid c_2 of p_2, moving p_1 and p_2 closer to one another.

F1 to F14 describe our force model that strives to generate improved layouts that satisfy objectives 1 to 6 (Section 4.2.3). We now describe the algorithm we used to apply and balance out the forces.

4.3.2 The Algorithm

Our algorithm is similar to that used by Eades [1984] to apply and balance out the forces in the system. Given an Euler diagram generated by a current automatic drawing method, **the** euler***Force*** **algorithm** iteratively: checks which of the forces in Section 4.3.1 should be applied to the layout of the diagram based on its current state and the set of zones in the diagram's initial layout; computes the magnitude and the direction of the forces on each of the structures (vertices, edges, entire polygons) that the forces should be applied to; moves these structures accordingly to create a new layout. After a number of iterations, the forces in the system equilibrate and the improved layout is returned.

Most of the forces in the model act on the vertices of the polygons in the layout, so polygons with fewer vertices are subject to fewer changes than those with more vertices. Thus, before the algorithms iterates through the force model, **the number of vertices on each of the polygons** in the given diagram layout is equalized. For instance, if a layout has two polygons, p_1 and p_2, and p_1 has 10 vertices and p_2 has 12 vertices, two vertices are added to p_1. This is done by first adding a vertex x between two vertices v_1 and v_2 of the polygon that are connected by an edge (v_1, v_2), then removing (v_1, v_2) and adding new edges (v_1, x) and (x, v_2). Since the forces can enlarge the polygons, at the end of every iteration, vertices are added on edges whose length is large enough to make the polygon look non-smooth.

Algorithm 4.1 shows **the** euler***Force*** **algorithm** as explained above. Figure 4.21 illustrates the layouts produced at various stages of the algorithm for a diagram example provided as an input to euler*Force*.

Algorithm 4.1: eulerForce (*d*)

Input: An Euler diagram *d* with curves represented as polygons
Output: An improved layout for *d*

1.   *layout* ← *d*
2.   Equalize the number of vertices of each polygon in *layout*
3.   **for** *i* = 1 to *I* **do**
4.   **for** each force *F* in the force model **do**
5.   **if** *F* should be applied on *layout* based on *layout*'s current state and the set of zones in *d* **then**
6.   **for** each structure *s* in *layout* on which *F* should be applied **do**
7.   (*m*, *r*) ← magnitude and direction of *F* on *s*
8.   *layout* ← new layout after *s* is moved *m* in direction *r*
9.   **end for**
10.   **end if**
11.   **end for**
12.   Add as many vertices as necessary to the polygons in *layout*, so the polygons look smooth
13. **end for**
14. **return** *layout*

For Euler diagrams with 3 curves *I*=200, 4-curves *I*=250, 5-curves *I*=350, more than 5 curves *I*=500.

A – *0 iterations*   **B** – *100 iterations*
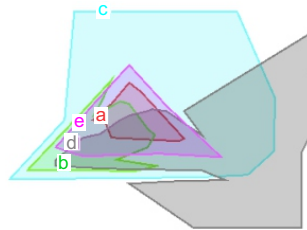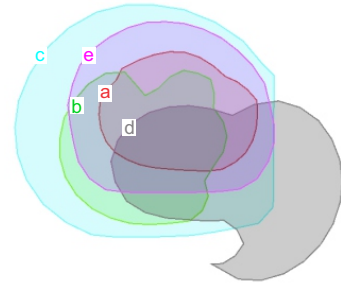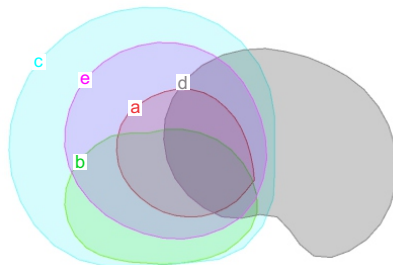
C – *200 iterations*   **D** – *350 iterations – final*
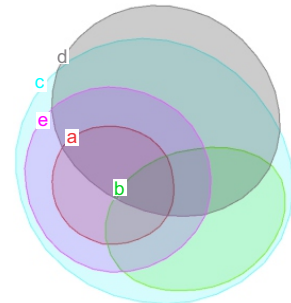
Figure 4.21: *Running the* eulerForce *algorithm on a diagram example.* (A) A diagram generated by Rodgers et al.'s [2008b] method with abstract description {∅, *c, d, bc, cd, ce, ace, bcd, bce, cde, abce, acde, bcde, abcde*} that is passed on as input to eulerForce. The layout of A after: (B) 100 iterations; (C) 200 iterations; (D) 350 iterations, when the algorithm terminates. D is the improved layout returned by eulerForce for A. Though B has a zone *de* which is not visible in A, the set of zones of C and D is the same as that of A.

The **number of iterations *I*** in Algorithm 4.1 is predefined (hardcoded), as in Eades's [1984] algorithm, but its value varies depending on the number of curves in the Euler diagram. The value of *I* for an *n*-Euler diagram, shown below Algorithm 4.1, was determined after the layout improvement process was observed for a number of different *n*-Euler diagrams using different values for *I*. The chosen values were the ones that best met our objectives in Section 4.2.3 for the tested diagrams.

Due to the various forces in the model, **a limit is set on the magnitude of the resultant force exerted on a structure** (vertex, edge, polygon). This limit is inversely proportional to the current *i* in Algorithm 4.1, so major changes are only carried out at the initial iterations when a more extensive search for an appropriate layout is required. During the final iterations, minor changes are carried out to refine the layout and ensure the algorithm converges to a solution.

The **transition from the initial to the final layout** is animated, thus facilitating understanding of how the forces aid in improving the layout and how they interact with one another [Di Battista et al., 1999a]. This method was also helpful to understand and appropriately define the required forces to lay out Euler diagrams and to devise the first force model to improve their layout. Moreover, this simple algorithm could aid in preserving the layout's mental map [Eades et al., 1991] despite it is changed.

Eades's [1984] **simple spring embedder was aimed for non-dense graphs with few vertices**. For graphs with with hundreds of vertices, poor layouts are produced by this embedder due to the various local minima in its physical model [Kobourov, 2012]. As discussed in Section 4.3.1, we avoid these local minima by accepting all layouts obtained during the improvement process, including those that do not depict the same set of zones as in the initial layout. To correct these layouts, we then apply forces F4, F6, F8, F13 and F14 to regain lost zones and remove extra zones. In this way, we explore more of the search space than is possible if we disallow incorrect layouts from the improvement process. Even so, Euler diagram layouts typically have few curves, thus fewer than hundreds of vertices. Sophisticated techniques, like those in Section 2.4.1, will later be adopted to improve the efficiency and performance of our algorithm. Example, Hu's [2005] method lays out graphs with over 10,000 vertices in less than a minute. Yet as shown in Section 4.4, eulerForce already produces improved layouts in relatively fast time.

4.4 Evaluation

To evaluate our method eulerForce, we used its software implementation to improve the layouts of Euler diagrams generated by a current drawing method (Section 4.4.1), and we compared eulerForce's layouts with those generated by the only other implemented layout method for Euler diagrams (Section 4.4.2). All the experiments were run on an Intel Core 2 Duo CPU E7200 @2.53GHz with 3.23GB RAM, 32-bit Microsoft Windows XP Professional SP1, SP2 and SP3 and Java Platform 1.6.0.14.

4.4.1 Accuracy, Time and Aesthetics

We tested eulerForce on diagrams automatically generated by Rodgers et al.'s [2008b] method, to evaluate its effectiveness in generating improved layouts that satisfy our objectives 1-6 (Section 4.2.3). Rodgers et al.'s method was chosen, as it is the only method that draws a diagram for abstract descriptions for which a wellformed Euler diagram can be drawn (Section 3.6). Thus, if an improved layout generated by eulerForce had disconnected zones or concurrent curves, such that it was non-wellformed and it did not satisfy our objective 3 or 4 (Section 4.2.3), a limitation in our method was evident, as a wellformed diagram for that abstract description is known to exist.
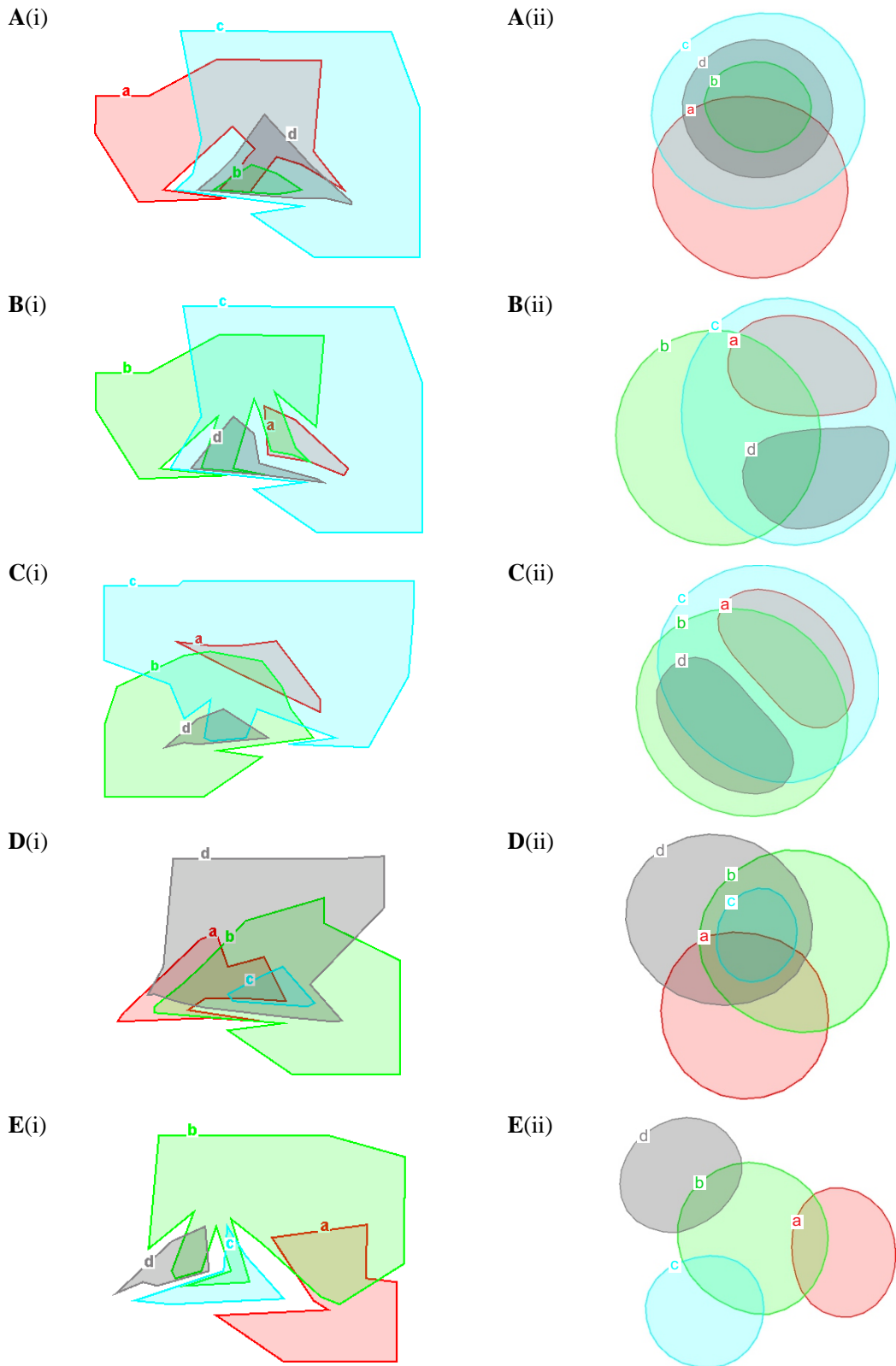
Figure 4.22: *Examples of diagrams in the library generated by Rodgers et al.'s [2008b] method (left, (i)) and* eulerForce*'s layout for each of the diagrams (right, (ii))—the diagrams have four curves and* eulerForce*'s layouts are correct, as each (ii) depicts the same set of zones as respective (i).* Layouts (i) and (ii) have abstract description: (A) {∅, *a*, *c*, *ac*, *cd*, *acd*, *bcd*, *abcd*}; (B) {∅, *b*, *c*, *ac*, *bc*, *cd*, *abc*, *bcd*}; (C) {∅, *b*, *c*, *ac*, *bc*, *bd*, *abc*, *bcd*}; (D) {∅, *a*, *b*, *d*, *ab*, *ad*, *bd*, *abd*, *bcd*, *abcd*}; (E) {∅, *a*, *b*, *c*, *d*, *ab*, *bc*, *bd*}.
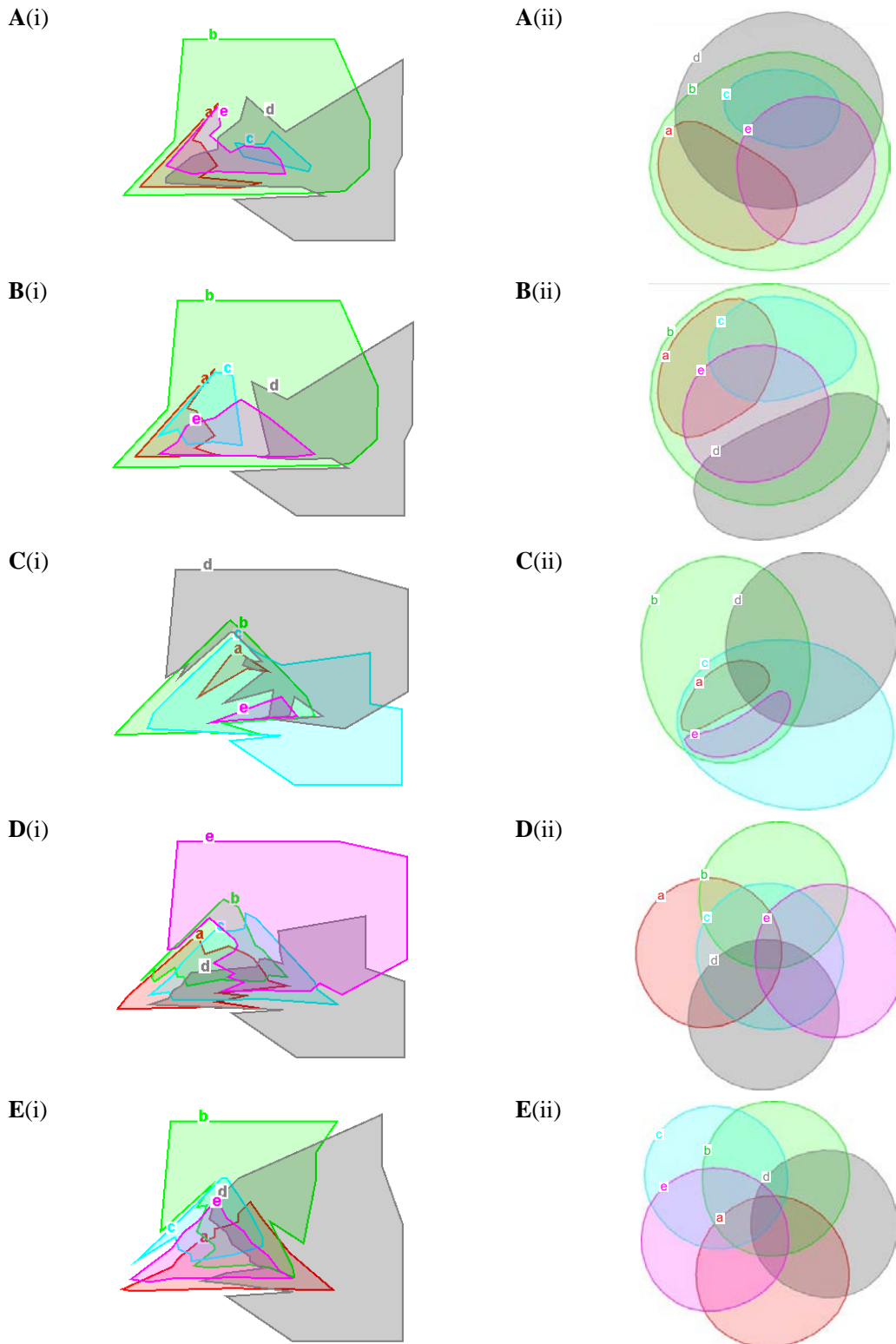
Figure 4.23: *Examples of diagrams in the library generated by Rodgers et al.'s [2008b] method (left, (i)) and* eulerForce*'s layout for each of the diagrams (right, (ii))—the diagrams have five curves and* eulerForce*'s layouts are correct, as each (ii) depicts the same set of zones as respective (i).* Layouts (i) and (ii) have abstract description: (A) {∅, *b*, *d*, *ab*, *bd*, *be*, *abd*, *abe*, *bcd*, *bde*, *abde*, *bcde*}; (B) {∅, *b*, *d*, *ab*, *bc*, *bd*, *be*, *abc*, *abe*, *bce*, *bde*, *abce*}; (C) {∅, *b*, *c*, *d*, *bc*, *bd*, *cd*, *abc*, *bcd*, *bce*, *abcd*, *bcde*}; (D) {∅, *a*, *b*, *d*, *e*, *ab*, *ac*, *ad*, *bc*, *be*, *cd*, *ce*, *de*, *abc*, *acd*, *bce*, *cde*, *abcd*, *abce*, *acde*, *bcde*, *abcde*}; (E) {∅, *a*, *b*, *c*, *d*, *e*, *ad*, *ae*, *bc*, *bd*, *ce*, *abd*, *abe*, *ace*, *ade*, *bcd*, *bce*, *abcd*, *abce*, *abde*, *bcde*, *abcde*}.
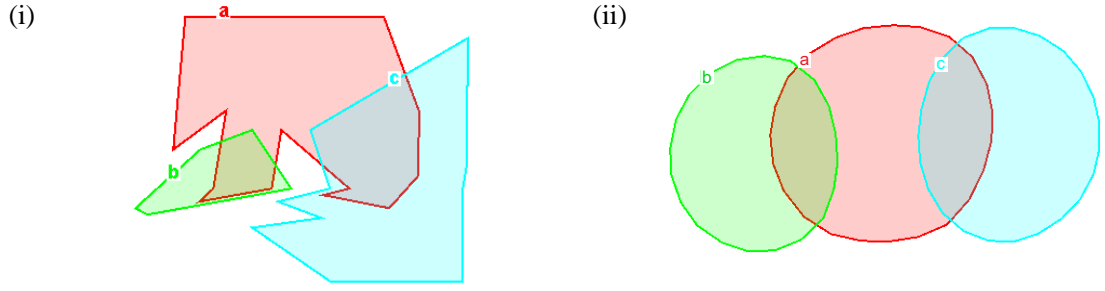
Figure 4.24: *Examples of diagrams in the library generated by Rodgers et al.'s [2008b] method (left, (i)) and* eulerForce*'s layout for each of the diagrams (right, (ii))—the diagram has three curves and* eulerForce*'s layout is correct, as (ii) depicts the same set of zones as (i).* Layouts (i) and (ii) have abstract description $\{\varnothing, a, b, c, ab, ac\}$.
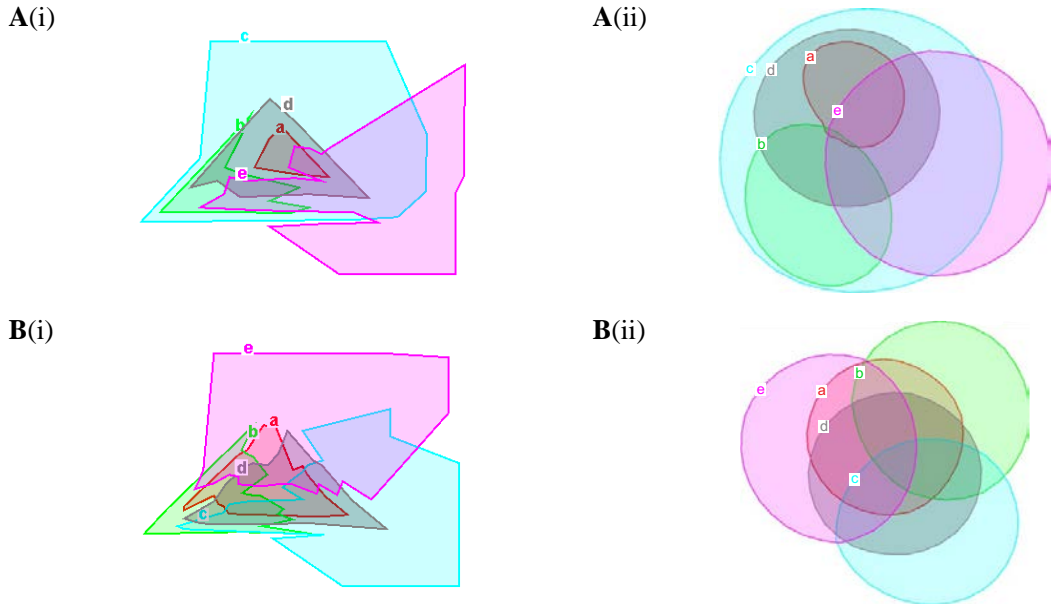
Figure 4.25: *Examples of diagrams in the library generated by Rodgers et al.'s [2008b] method (left, (i)) and* eulerForce*'s layout for each of the diagrams (right, (ii))—the diagrams have five curves and* eulerForce*'s layouts are incorrect, as each (ii) depicts a different set of zones than respective (i).* (A) The abstract description of (i) is $\{\varnothing, c, e, bc, cd, ce, acd, bcd, bce, cde, acde, bcde\}$, but that of (ii) is $\{\varnothing, c, e, bc, cd, ce, acd, bcd, bce, cde, acde, bcde, abcd, abcde\}$ as it has zones, *abcd* and *abcde*, that are not depicted in (i). Also, curve *a* in (ii) is not completely regular and smooth. (B) The abstract description of (i) is $\{\varnothing, b, c, e, ab, ad, ae, bc, bd, be, cd, ce, de, abd, abe, acd, ade, bcd, cde, abcd, abde, acde\}$, but that of (ii) is $\{\varnothing, b, c, e, ab, ae, bc, bd, cd, ce, de, abd, abe, acd, ade, bcd, cde, abcd, abde, acde, abcde\}$ as it does not have zones, *ad* and *be*, that are depicted in (i) but has zone, *abcde*, that is not depicted in (i). Also, (ii) is non-wellformed.

A library of Euler diagrams generated by Rodgers et al.'s [2008b] method for all the abstract descriptions for which a wellformed Euler diagram with three, four and five curves can be drawn was assembled. This library included: 9 3-Euler diagrams, 114 4-Euler diagrams, 342 5-Euler diagrams. Our method euler**Force** **was used to improve the layout of the diagrams in this library**. Figure 4.22, Figure 4.23, Figure 4.24 and Figure 4.25 illustrate a few of: (i) the diagrams in the library, (ii) the corresponding layout by eulerForce. Layouts (ii) in Figure 4.22, Figure 4.23, Figure 4.24 depict the same set of zones as those in (i), but those in Figure 4.25 do not and are thus examples of cases where eulerForce fails to produce an appropriate layout. We now discuss eulerForce's results shown in Table 4.1.

Table 4.1: *Results for the layouts produced by* euler*Force for all of Rodgers et al.'s [2008b] diagrams in our library: the percentage of layouts satisfying our objectives; the mean time to produce the layouts.*

		objectives						
	N	1 regular, smooth, similarly shaped, convex curves	2 same zones as initial layout	3 connected zones	4 curves not too close to one another	5 centring contained curves in containing curve or zone	6 adequately sized zones and curves	mean time
3-Euler	9	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	7s
4-Euler	114	94.4%	100.0%	100.0%	93.8%	99.1%	72.8%	26s
5-Euler	342	92.4%	61.2%	84.5%	88.0%	96.5%	58.1%	77s

An *n*-Euler is an Euler diagram with *n* curves. Our objectives were defined in Section 4.2.3.

Accuracy

A layout produced by euler*Force* is ***accurate* if it satisfies objective 2** (Section 4.2.3) and thus depicts the same set of zones as in the initial layout. Table 4.1 shows that all the produced layouts were accurate except for those with five curves. Most (95.5%) of the **inaccurate 5-curve layouts** had on average 200 vertices and were produced for dense Euler diagrams with over 20 zones. Spring embedders are aimed for non-dense graphs with less than 50 vertices [Eades, 1984; Gibson et al., 2012], so this result was partly expected as spring embedders cannot effectively manage the forces on a large number of structures. Yet, as discussed in Section 4.2.3 and Section 4.3.2, we opted for a spring embedder to help us identify (for the first time) the main forces that are required to handle Euler diagrams. Sophisticated force-directed techniques (Section 2.4.1) will later be used to better handle large, dense Euler diagrams.

In Figure 4.25A, euler*Force*'s layout (ii) **has zones** *abcd* and *abcde* that are **not depicted in the initial diagram layout** (i), as curves *a* and *b* should be disjoint as in Figure 4.26A. The non-smooth section of curve *a* indicates that force F4 was attempting to get curves *a* and *b* disjoint, but did not fully succeed as F4 was weaker than other current interacting forces.

In Figure 4.25B, euler*Force*'s layout (ii): **does not have zones** *ad* and *be* that are **depicted in the initial layout** (i); **has zone** *abcde* that is **not depicted in the initial layout** (i). Despite having smooth curves, layout (ii) is non-wellformed due to a triple point on curves *a*, *b* and *e*. To depict multiple curve overlaps, less regular shaped curves are likely to be required and thus, the strength of forces F1 and F2 (to obtain regular, smooth, similarly shaped, convex curves—objective 1) might have to be dynamically tuned using more sophisticated techniques. In fact, to depict the set of zones of (i) with smooth curves and a wellformed layout, curves *b*, *c*, *e* should be less circular and more elongated as in Figure 4.26B.
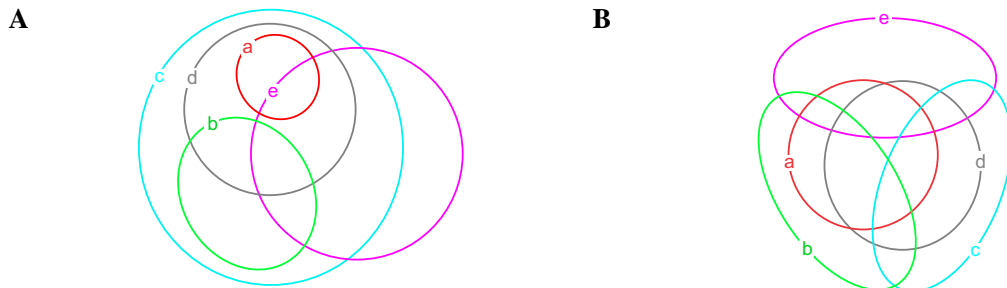
Figure 4.26: *Alternative layouts to Figure 4.25A(i) and Figure 4.25B(i).* (A) Has abstract description {∅, *c*, *e*, *bc*, *cd*, *ce*, *acd*, *bcd*, *bce*, *cde*, *acde*, *bcde*} as Figure 4.25A(i). (B) Has abstract description {∅, *b*, *c*, *e*, *ab*, *ad*, *ae*, *bc*, *bd*, *be*, *cd*, *ce*, *de*, *abd*, *abe*, *acd*, *ade*, *bcd*, *cde*, *abcd*, *abde*, *acde*} as Figure 4.25B(i).

Time

Despite using a simple algorithm, eulerForce still **produces layouts in relatively fast time** (Table 4.1). This time is comparable to that of other force-directed drawing algorithms for graphs, which typically produce layouts within a minute [Gibson et al., 2012]. Also, a response time of 10 seconds or less ensures the users' attention is maintained [Miller, 1968]. This time could be improved further in the future once a better-optimized algorithm (like those in Section 2.4.1) is adopted.

Aesthetics

Besides being accurate, eulerForce's layouts should **satisfy the aesthetic criteria in objectives 1, 3-6** (Section 4.2.3). Table 4.1 shows the percentage of eulerForce's layouts satisfying each of these objectives. No formal metrics are available to measure the aesthetics of Euler diagrams (Section 3.5) and thus, the results in Table 4.1 for objectives 1, 4, 5 are based on our judgement of how these should be measured.

A layout with **regular, smooth, similarly shaped, convex curves** (objective 1) has: polygons for the curves whose interior angles are greater than $3\pi/4$; curves that can be classified into no more than two groups of similarly shaped curves (e.g., in Figure 4.22C(ii), the two groups are: curves b and c; curves a and d). There are no formal metrics for this aesthetic criteria and thus, our measure is based on our judgement of what constitutes a smooth curve and a layout with similarly shaped curves. Table 4.1 indicates that most of the layouts, including inaccurate ones (e.g., Figure 4.25B(ii)) satisfied objective 1, which is typical for a spring embedder due to its physical system and its attractive and repulsive forces. The same is apparent in (ii) of Figure 4.22, Figure 4.23, Figure 4.24 and Figure 4.25.

A layout with **connected zones** (objective 3) has zones with only one minimal region (Section 3.4.2). All accurate layouts were wellformed, thus all satisfied objective 3, as in (ii) of Figure 4.22, Figure 4.23, Figure 4.24. Only 39.8% of the inaccurate layouts, all of which had five curves, had disconnected zones.

A layout has **curves that are not too close to one another** (objective 4) when all of its non-intersecting curves are at least 10 pixels away from one another. There are no formal metrics for this aesthetic criteria and thus, our measure is based on our judgement of how far apart two non-intersecting curves have to be for them to be seen as discrete and non-concurrent. All accurate layouts satisfied objective 4, as in (ii) of Figure 4.22, Figure 4.23, Figure 4.24. Only 16.5% of the inaccurate layouts, all of which had five curves, did not satisfy objective 4.

A layout whose **contained curves are centred in their containing curve or zone** (objective 5) is one whose curves contained in other curves or zones appear to be centred as appropriately as possible in their containing curve or zone. It is difficult to formally evaluate how well this objective was met as the possibility of centring a curve depends on the layout of other curves and the required zones. Example in (ii) of Figure 4.23C, curve a cannot be centred in the containing zone bc, as zone bc also contains curve e. Table 4.1 shows that most contained curves appeared to be (according to our visual judgement) centred as appropriately as possible. Layouts (ii) of Figure 4.22A-D and Figure 4.23A-C indicate that the force (F10) that handles this objective aids in improving the aesthetics of the layouts as it ensures that: contained curves are evenly distributed in the containing curve or zone, thus helping produce highly symmetric layouts; containing curves are not enlarged unnecessarily just because a contained curve is getting too close to its edge despite that there is enough space elsewhere in the containing zone.

A layout with **adequately sized zones and curves** (objective 6) is one whose area for each of the zones is inversely proportional to the number of curves in which the zone is located (Section 4.3.1).

Most of the layouts satisfied objective 6 (Table 4.1), but as expected, fewer layouts satisfied this objective than others as the forces (F11-F14) handling this objective were intentionally weaker than others that handle more important aesthetics. Layouts (ii) of Figure 4.22, Figure 4.23 and Figure 4.24 indicate that our approach to ensure adequately sized zones and curves is effective as it ensures that: curves contained in other curves or zones are not too large to fit in the containing curve or zone with possibly other zones as in Figure 4.22B and Figure 4.23A-C and without breaking wellformedness; the number of curves in which a zone is located is easier to identify than if all the zones had a similar area.

Layouts (ii) of Figure 4.22, Figure 4.23, Figure 4.24 indicate that eulerForce's **layouts are often symmetric**. The typical spring embedder forces (F1, F2) are likely to produce symmetric layouts [Eades and Lin, 2000]. Yet other forces in our model like the centring force F10 help in achieving symmetry.

For the layouts to be effectively evaluated, formalized aesthetic metrics and cognitive measures (like those in graph drawing [Purchase et al., 2002, 2002, 2012]; Appendix C.1.1) are required. Few studies have investigated the aesthetics of such diagrams and **no criteria have been formalized** (Section 3.5).

4.4.2 eulerForce versus Previous Methods

The only other implemented layout method is **Flower et al.'s [2003b] multi-criteria optimization method** (Section 4.2.1). We visually compared layouts generated by eulerForce and Flower et al.'s method ((ii) in Figure 4.27 and Figure 4.28) for two 4-set abstract descriptions (Flower et al.'s method handles up to four curves). Table 4.2 shows our observations. There are no formal metrics or empirical evidence for Euler diagram aesthetics (Section 3.5) and thus, our judgement could be partly subjective.

Flower et al.'s **initial layouts** were generated by methods [Flower et al., 2003a; Flower and Howse, 2002] available at the time. The initial layouts of eulerForce were generated by a more recent method [Rodgers et al., 2008b], which is a variant of those used by Flower et al., as it uses a planar dual graph to draw the curves and it is restricted to abstract descriptions for which a wellformed Euler diagram can be drawn (Section 3.6). So for the same abstract description, the initial layout used by Flower et al.'s method (e.g., (i) of Figure 4.27) was different from that used by eulerForce (e.g., (i) of Figure 4.28).

For abstract description {∅, *a*, *c*, *ac*, *cd*, *acd*, *bcd*, *abcd*}, Flower et al.'s method produced layout (ii) from the initial layout (i) in Figure 4.27A, while eulerForce produced layout (ii) from the initial layout (i) in Figure 4.28A. The layouts by both methods are accurate as they depict the same zones as in the initial layout. The curves in Flower et al.'s layout are smooth and convex, but different in shape, and the layout is non-symmetric. The curves in eulerForce's layout have a more regular, circular, similar shape than those in Flower et al.'s layout. These curve aesthetics are preferred and consistent with various perceptual principles that could facilitate diagram understanding (Section 3.5.2). The containing and contained curves *b, c* and *d* are centre aligned and the distance between curve *c* and *d* is the same as the distance between curve *b* and *d*. These features could make subsets more apparent and the layout more symmetric. It is unknown whether symmetry could aid Euler diagram comprehension, but it is known that symmetry helps comprehension of other diagrams like graphs (e.g., [Purchase et al., 2002]).

For abstract description {∅, *a*, *c*, *d*, *ac*, *ad*, *bc*, *abc*}, Flower et al.'s method produced the accurate layout (ii) from the initial layout (i) in Figure 4.27B, while eulerForce produced the accurate layout (ii) from the initial layout (i) in Figure 4.28B. The curve aesthetics of the layouts produced by Flower et al. and eulerForce are similar to those of the layouts in the previous example (Figure 4.27A, Figure 4.28A).
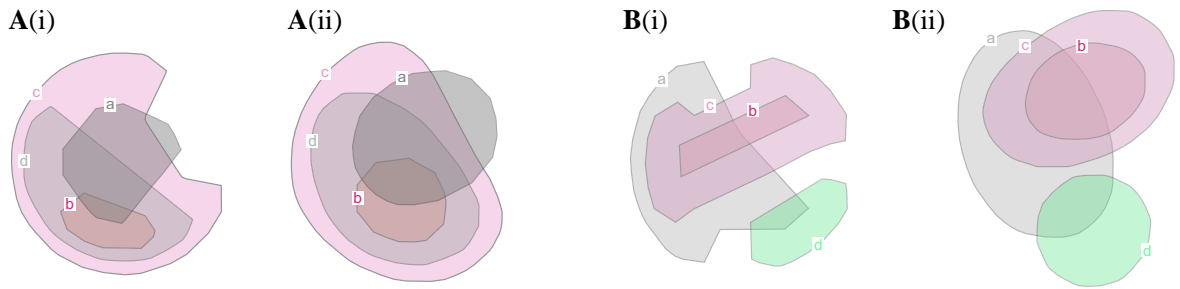
Figure 4.27: *The improved layouts (ii) generated by Flower et al.'s [2003b] method for the diagrams and initial layouts (i).* Curve labels were added to the original diagrams in Flower et al.'s [2003b] article. Layouts (i) and (ii) have abstract descriptions (A) {∅, *a*, *c*, *ac*, *cd*, *acd*, *bcd*, *abcd*} and (B) {∅, *a*, *c*, *d*, *ac*, *ad*, *bc*, *abc*}, which are the same as those in Figure 4.28. *Source*: [Flower et al., 2003b] Figures 11 and 13 redrawn
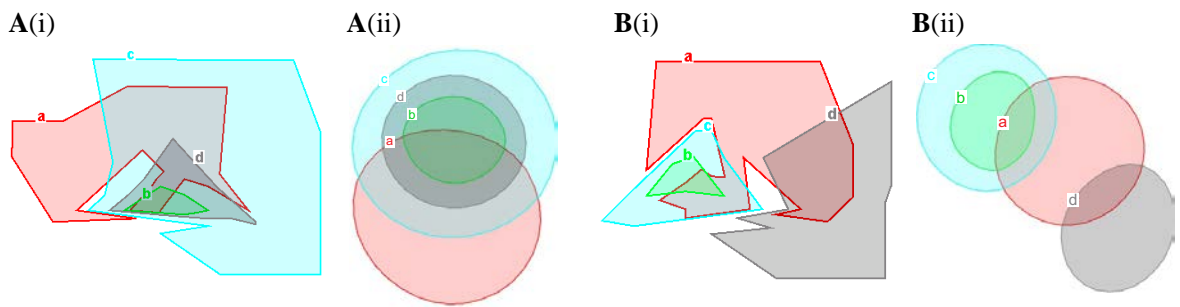
Figure 4.28: *The improved layouts (ii) generated by* eulerForce *for the diagrams and initial layouts (i).* Layouts (i) and (ii) have abstract descriptions (A) {∅, *a*, *c*, *ac*, *cd*, *acd*, *bcd*, *abcd*} and (B) {∅, *a*, *c*, *d*, *ac*, *ad*, *bc*, *abc*}, which are the same as those in Figure 4.27.

Table 4.2: *The characteristics of the two layouts (ii) in Figure 4.27 produced by Flower et al.'s [2003b] method and the characteristics of the two layouts (ii) in Figure 4.28 produced by* eulerForce.

	same zones as initial layout	smooth curves	convex curves	similarly shaped curves	contained curves centred in containing curve	symmetric layout	adequately sized zones and curves
Flower et al.	yes	yes	yes	somewhat	somewhat	no	no
eulerForce	yes	yes	yes	yes	yes	yes	yes

yes / somewhat / no – the 2 layouts generated by the method *had / somewhat had / did not have* the characteristic

Both eulerForce and Flower et al.'s method aim for **layouts with adequately sized zones**. Our method eulerForce aims for a zone area that is inversely proportional to the number of curves in which the zone is located, while Flower et al.'s method aims for same sized zones. As discussed in Section 4.3.1, the latter definition for an adequate zone area is not appropriate for all Euler diagrams. In fact, Flower et al. indicate that often their optimization is not able to optimize the zone areas, as in most cases same sized zones cannot be obtained without breaking other more important aesthetic criteria like curve roundness and closeness. Thus, the zones often end up getting an arbitrary non-optimized area. This is evident in Figure 4.27, as Flower et al.'s produced layouts seem to have differently sized zones. In contrast, all the zones located in the same number of curves of eulerForce's layouts in Figure 4.28 seem to have a similar area. As discussed in Section 4.3.1 and as noted by Flower et al., it is hard to get same sized zones in an accurate layout that satisfies important curves aesthetics. Thus other definitions of adequately sized zones, like the one adopted by eulerForce, should be considered and evaluated.

Though the initial layouts used by euler*Force* have curves that are less smooth and less convex than those used by Flower et al.'s method, euler**Force's improved layouts** are accurate and have smooth, convex curves as in Flower et al.'s layouts, but they also have similarly shaped curves, contained curves centred in containing curves, a symmetric layout, and adequately (non-arbitrary) sized zones, all of which are not as apparent in Flower et al.'s layouts (Table 4.2).

4.5 Future Work

This is the first forced-directed approach that has been devised to improve the aesthetics of the curves of Euler diagrams and though we adopted a simple spring embedder, our evaluation (Section 4.4) indicates a great potential for this approach. So further investigations should be carried out to improve and extend our method, so the layout of different diagram types is improved in relatively fast time.

Firstly, it would be interesting to evaluate the layouts generated by euler*Force* for initial layouts that are non-wellformed and for diagrams with an abstract description for which a wellformed Euler diagram cannot be drawn. Until now, euler*Force* has been evaluated for initial layouts that are wellformed and have an abstract description for which a wellformed diagram can be drawn (Section 4.4.1). This was intentionally done to evaluate the effectiveness of the forces that we specifically devised to ensure that the zones are connected and the curves are not too close to one another (objectives 3 and 4, Section 4.2.3; forces described in Section 4.3.1 subsection 'Forces to Meet Objective 3' and subsection 'Forces to Meet Objective 4'). However, the effectiveness of these forces in handling non-wellformed diagrams should be evaluated, so that, if necessary, the force model is adapted to handle such diagrams.

Our main aim was to devise a force model that appropriately handles Euler diagrams. We wanted to evaluate the effectiveness of our forces and how they interact with one another and thus, we opted for Eades's [1984] simple spring embedder algorithm that facilitates understanding of the interacting forces in the system [Di Battista et al., 1999a]. We are aware that this algorithm is not as efficient as other algorithms [Gibson et al., 2012] and is unable to appropriately handle hundreds of vertices due to the likelihood of reaching a local minimum [Kobourov, 2012]. Such limitations are evident in our euler*Force* evaluation for Euler diagram layout with five curves (discussed in Section 4.4.1 subsection 'Accuracy'). However, until now, our focus was on the force model rather than the algorithm. Now that the force model is devised, sophisticated force-directed algorithms such as those used for laying out large graphs (discussed in Section 2.4.1) should be adopted and investigated.

For instance, a multilevel approach such as that used in graph drawing (e.g., [Hu, 2005; Walshaw, 2001]) can be adopted to overcome local minima and to efficiently handle layouts with thousands of vertices and thus, with various curves and zones like those in Figure 4.25. As an example, Hu's [2005] method uses this approach to lay out graphs with over 10,000 vertices in less than a minute.

Barnes and Hut [1986] algorithm can be used to efficiently and dynamically compute the appropriate forces at every step of the layout improvement process. This could thus aid in handling cases such as those in Figure 4.25 . This method has been successful in graph drawing to ensure the required criteria are met (e.g., [Hu, 2005; Tunkelang, 1999]). Force-directed techniques in graph drawing have also demonstrated that adding magnetic fields to the system and its springs could aid in satisfying various aesthetic criteria [Sugiyama and Misue, 1995a, b] and so, this technique should also be considering for Euler diagram layouts.

Very few studies have been carried out to assess the aesthetics of Euler diagrams and to identify the important features that make the diagrams easier to comprehend and reason out (e.g., [Benoy and Rodgers, 2007; Rodgers et al., 2012b]; Section 3.5). However, formalized aesthetic metrics and cognitive measures are required to appropriately evaluate Euler diagram layouts and their effectiveness. Though Flower et al. [2003b] defined a few metrics for their layout method, their metrics have not been evaluated and they might not be entirely appropriate in identifying effective layouts (Section 4.4.2). So at present, no formalized aesthetic criteria or metrics are available for Euler diagrams. As evident in graph drawing (graph aesthetics discussed in Appendix C.1.1), various studies, expertise, resources and time are required for such metrics and measures to be fully formalized and evaluated, but work should commence to ensure the availability of guidelines to appropriately evaluate Euler diagram layouts.

4.6 Summary

Drawing a comprehensible Euler diagram that only depicts the required zones in an abstract description is difficult (Section 3.6). Automatic drawing techniques that attempt to generate comprehensible Euler diagrams by for instance, restricting the shape of the curves to a circle, often generate diagrams with empty zones which could be misleading (Section 3.5.2). Other techniques that generate diagrams with precisely the required set of zones are often incomprehensible as the curves are non-smooth, non-convex and indistinguishable (Section 3.5.2). Two methods have been proposed to improve the layouts of such diagrams (Section 3.6.3). However, these methods are complex, computational expensive and inappropriate to satisfy diverse aesthetic criteria and handle diagrams with various curves. Force-directed techniques (Section 2.4.1) have been widely evaluated and used to lay out comprehensible graphs in relatively fast time, but have never been used for Euler diagram layouts.

In this chapter, we introduced our layout method, eulerForce, as the first method that uses a force-directed approach to improve the layout of Euler diagrams in relatively fast time. We discussed the force model we devised to handle Euler diagrams and our evaluation of this technique. Though this was a preliminary project and a simple algorithm was adopted, our evaluation indicates great potential for using force-directed techniques to improve Euler diagram layouts and to possibly develop an efficient automatic drawing technique that generates comprehensible diagrams given an abstract description.

In most application areas, 3-Venn diagrams that are area-proportional are typically used (Section 3.7.1), as in one diagram both the data set relations and cardinalities are depicted, thus facilitating analysis. In contrast to the diagrams handled in this chapter, in such cases, the area of each zone must be proportional to the cardinality of the depicted data set relation. So special drawing algorithms, different from those used for not area-proportional diagrams, are required to accurately generate such diagrams (Section 3.7.4). Typically, the curves of these diagrams are drawn as circles due to their regularity and good continuity. However, these diagrams are often inaccurate and misleading as circles have limited degrees of freedom (i.e., a radius and a centre). So polygons are used to produce accurate diagrams, but due to their non-smoothness, the diagrams are incomprehensible. Ellipses are smooth and have more degrees of freedom than circles (i.e., a centre, a semi-minor and a semi-major axis, and an angle of rotation) and so, they are more likely to produce diagrams that are both accurate and easy to comprehend. However, none of the current drawing techniques, before our method eulerAPE, use ellipses. In the next chapter, we discuss eulerAPE and the effectiveness of ellipses in drawing accurate and comprehensible area-proportional Venn diagrams with three curves.

Chapter 5

Drawing Area-Proportional 3-Venn Diagrams Using Ellipses

This chapter introduces our drawing algorithm eulerAPE *for area-proportional 3-Venn diagrams using ellipses. Its software implementation at* `http://www.eulerdiagrams.org/eulerAPE`. *It also includes: analytic algorithms that compute the region areas of three overlapping ellipses, an evaluation of the effectiveness of ellipses in drawing area-proportional 3-Venn diagrams, characteristics of the area specifications that ellipses can depict accurately, and diagram design features to aid comprehension.*

5.1 Introduction

Area-proportional 3-Venn diagrams (Section 3.7) are used extensively in diverse areas (Section 3.7.1) to visualize data set relationships and cardinalities. Drawing these diagrams manually is difficult and typically, an automatic drawing method (Section 3.7.4) is used. The data sets are often depicted as circles, as they perceptually pop out as complete distinct objects due to their regularity and good continuity (Section 3.5.2). However, circles cannot draw accurate area-proportional 3-Venn diagrams for most data (Section 3.7.6) and so, the generated diagrams often have misleading zone areas. Accurate diagrams can be drawn with polygons, but due to their non-smoothness and lack of good continuity, the curves are not easily distinguishable and these diagrams are difficult to comprehend (Section 3.5.2).

In this chapter, we present our drawing algorithm eulerAPE as the first automatic method that uses ellipses to draw area-proportional 3-Venn diagrams. Ellipses have more degrees of freedom than circles, and unlike polygons, ellipses are smooth. So area-proportional Venn diagrams drawn with ellipses often have distinguishable curves and zone areas that are proportional to the set intersection cardinalities.

Our evaluation of eulerAPE and ellipses indicated that for a large majority of random area specifications (86%, $N = 10000$), an accurate area-proportional 3-Venn diagram can be drawn with ellipses, even though there are area specifications for which such a diagram cannot be drawn accurately using convex curves (Section 3.7.5). This indicates great potential for using curves that are regular and smooth like circles, but more general and flexible.

We start by outlining the motivation and objectives of this work (Section 5.2), followed by basic concepts and definitions namely related to ellipses (Section 5.3). We then discuss our analytic methods to calculate the region areas of three general overlapping ellipses (Section 5.4) and we explain how we generate libraries of random area specifications for which it is known or yet uncertain that an accurate area-proportional 3-Venn diagram can be drawn with ellipses (Section 5.5). We proceed with a detailed explanation of our drawing method eulerAPE and a justification for our design decisions (Section 5.6), followed by our evaluation of the effectiveness of eulerAPE and ellipses in depicting accurate area-proportional 3-Venn diagrams with smooth curves for various random area specifications (Section 5.7). We then discuss diagram design features that could facilitate comprehension of the diagram and the depicted set relations and cardinalities (Section 5.8). We end this chapter with a discussion of possible future work (Section 5.9) and a summary of our contributions and this chapter (Section 5.10).

This work has been published as a journal article [Micallef and Rodgers, 2014a], workshop paper [Micallef and Rodgers, 2014c] and poster [Micallef and Rodgers, 2012], and presented at the 2012 Annual Grace Hopper Celebration (GHC) and ACM Student Research Competition (SRC).

The software implementation is at `http://www.eulerdiagrams.org/eulerAPE`. eulerAPE is widely used in various research and industry areas (e.g., medicine, genomics, biophysics, astrophysics, mind sciences, marketing), and it is recommended on PNNL's Venn Diagram Plotter webpage [Littlefield and Monroe, 2013] indicating that eulerAPE produces more appropriate diagrams than their own circle-based method. Diagrams drawn by eulerAPE are being included in various journal (e.g., [Burdon et al., 2013; Flowers et al., 2013; Gresnigt et al., 2013]) and newspaper (e.g., [Bloch and Fairfield, 2013]) articles.

5.2 Motivations and Objectives

Area-proportional Venn diagrams with two and three curves are most commonly used (Section 3.7.1). Automatic drawing methods have been proposed (Section 3.7.4), but they have a number of limitations.

5.2.1 Limitations of Previous Drawing Methods

Area-proportional Venn diagrams cannot be drawn analytically using a specific curve, so heuristics or numerical methods are used (Section 3.7.6). Often these diagrams are drawn using circles (Section 3.7.1) due to their regularity and good continuity (Section 3.5.2). Circles can draw accurate area-proportional Venn diagrams for any data with two sets, but not three (Section 3.7.6), so most of those with three circles are misleading, like those discussed in Section 3.7.6, which we show again in Figure 5.1. Other diagrams with circles do not depict all the required zones or are non-wellformed (Section 3.7.4).

Such misleading diagrams are generated by all drawing methods using circles, including venneuler [Wilkinson, 2012], which is considered the current best method in drawing accurate area-proportional diagrams with circles (illustrated in Figure 3.28). Also, venneuler's loss function *stress*, that is used by venneuler's optimization method to obtain the required diagram, could be relatively low even when the diagram is missing required zones (e.g., Figure 3.28B and Figure 3.28C). Diagrams with missing zones do not depict all the required set relations of the data and are thus misleading. However, since the diagram's *stress* value is low, venneuler's optimization method considers such inappropriate diagrams a good solution. So venneuler's statistical loss function *stress* might not be as appropriate in ensuring that the required set relations are depicted.

Polygons have more degrees of freedom than circles and can thus draw more accurate diagrams. Chow proved that accurate, wellformed area-proportional 3-Venn diagrams with respect to any area specification can be drawn using orthogonal rectangles or 6-gons (Section 3.7.5). However, polygons are non-smooth, so the curves and zones are not easily distinguishable and the generated diagrams are difficult to comprehend, like those discussed in Section 3.7.6, which we show again in Figure 5.2. Thus, very few diagrams are drawn using polygons (Section 3.7.1).

Yet, drawing techniques, prior to our work in this chapter, depicted the curves as circles or polygons.
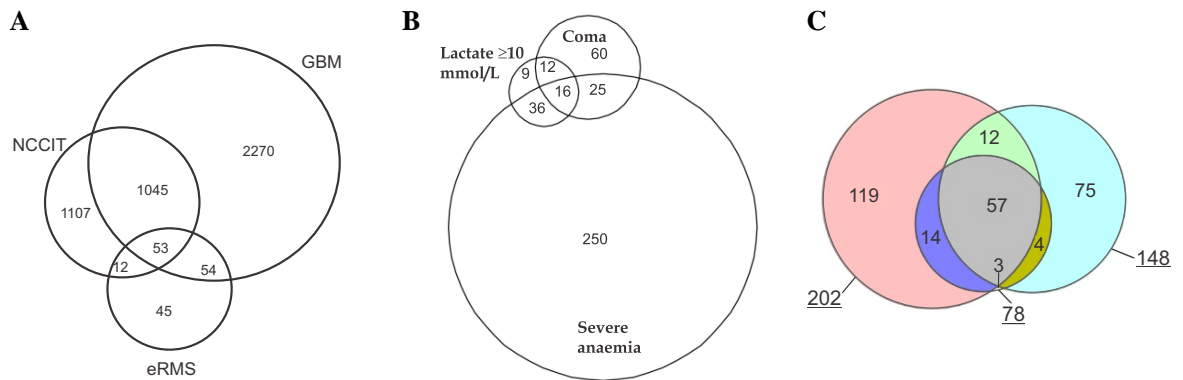
Figure 5.1: *Area-proportional 3-Venn diagrams drawn with circles in literature.* Figure 3.20 replica
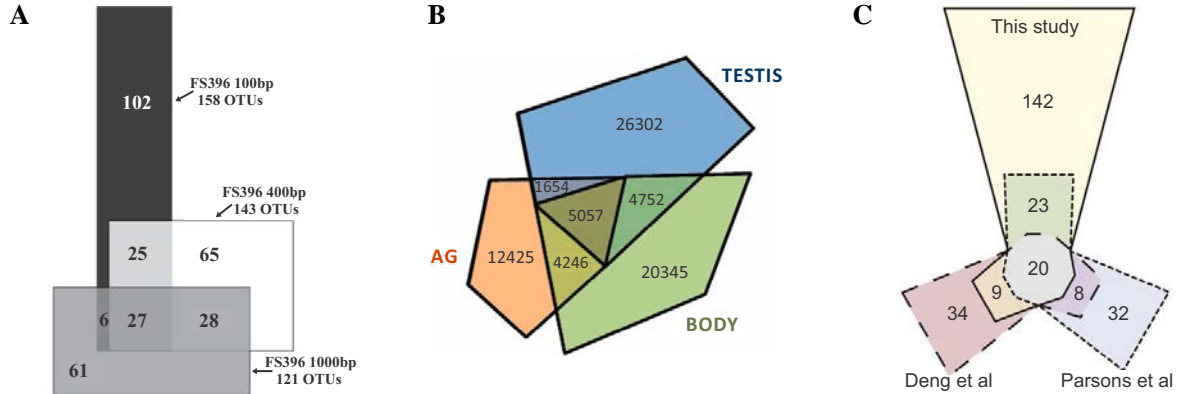
Figure 5.2: *Area-proportional 3-Venn diagrams drawn with polygons in literature.* Figure 3.21 replica

5.2.2 Using Curves with More Degrees of Freedom That Are Still Smooth

Using ellipses, the accurate diagrams with smooth, distinguishable curves in Figure 5.3 and Figure 5.4 were generated for the quantitative data indicated by the numeric labels of the respective diagram in Figure 5.1 or Figure 5.2. The diagrams in Figure 5.3 and Figure 5.4 were drawn using our novel drawing algorithm euler*APE*, discussed in Section 5.6. It is the first to use ellipses, so the few area-proportional 3-Venn diagrams with ellipses available prior to euler*APE* (e.g., [Gookin et al., 2008; Haetzman et al., 2003; Zhang et al., 2011]) were all drawn manually. The need to use ellipses was noted in 2004 in the first paper on drawing area-proportional Venn and Euler diagrams [Chow and Ruskey, 2004], and then in 2007 [Chow, 2007] and 2012 [Wilkinson, 2012]. Even so, ellipses had never been used due to difficulties in calculating the areas of the regions of overlapping ellipses and in adjusting the various properties of the ellipses [Chow and Ruskey, 2004].
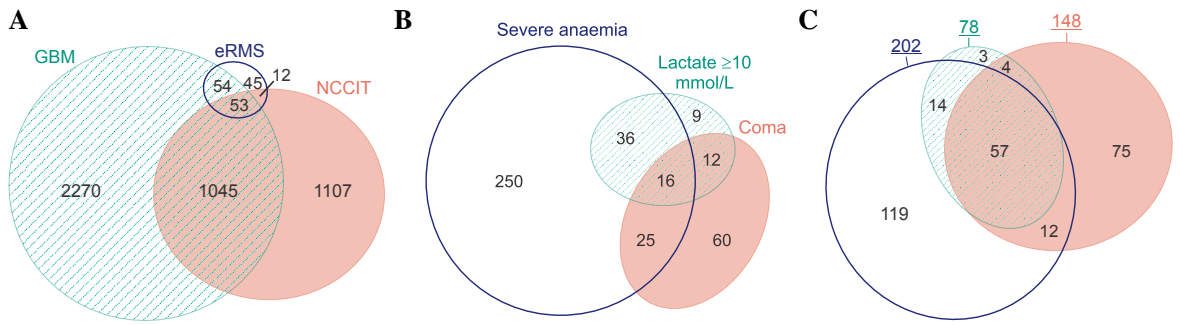
Figure 5.3: *Accurate area-proportional 3-Venn diagrams drawn with ellipses using our drawing algorithm* eulerAPE *for the quantities indicated by the numeric labels of the diagrams in Figure 5.1.*
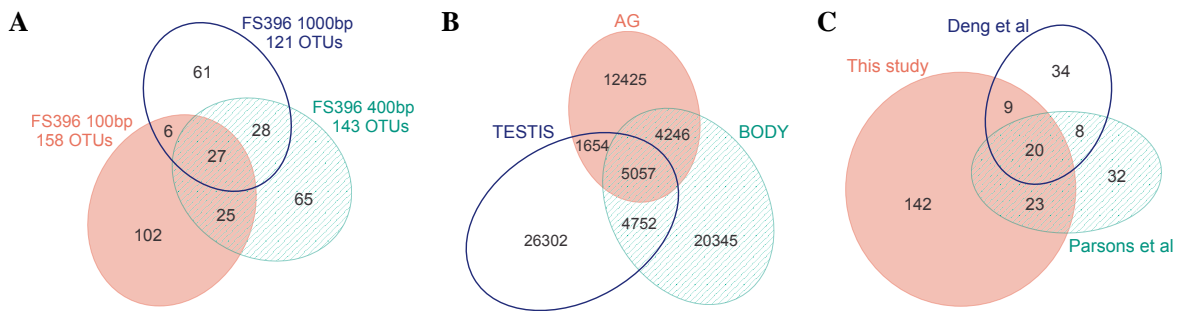
Figure 5.4: *Accurate area-proportional 3-Venn diagrams drawn with ellipses using our drawing algorithm* eulerAPE *for the quantities indicated by the numeric labels of the diagrams in Figure 5.2.*

There are area specifications for which an accurate area-proportional wellformed 3-Venn diagram cannot be drawn using convex curves (Section 3.7.5). So ellipses cannot generate accurate diagrams for all possible area specifications. However, since circles are preferred and a circle is a special instance of an ellipse, it is sensible to use ellipses whenever circles are not good enough. It would be beneficial to consider different shaped curves in a "natural progression" [Chow, 2007, p. 83], starting off with circles and progressing with more general shapes like ellipses and ovals before less desired and non-smooth shapes like polygons are considered. This would ensure that the smoothest and most regular curves possible for the required data are used. Thus ellipses are the next to be investigated.

5.2.3 Our Objectives

With the development of eulerAPE, we aspired to:

1. provide a deterministic drawing method that automatically draws appropriate area-proportional Venn diagrams with smooth curves for most data with three sets instantaneously;

2. evaluate the effectiveness of ellipses in drawing appropriate area-proportional Venn diagrams for most data with three sets, in contrast to circles and polygons;

3. provide and discuss diagram design features that could facilitate understanding of the diagram and the depicted set relations and cardinalities.

To achieve these objectives, we first devised eulerAPE (objective 1). Our method eulerAPE uses a simple hill-climber as an optimization technique to find a diagram that appropriately depicts the required set relations and cardinalities and so, the following components were required:

- analytic geometry methods to compute the region areas of three general intersecting ellipses (Section 5.4);

- a novel cost function to direct the optimization process towards a good solution (Section 5.6.4);

- a method to generate a rational starting diagram for the optimization (Section 5.6.5);

- a mechanism to adjust the properties of the ellipses during the optimization process in search for a good solution (Section 5.6.6).

There are various ways how each of the above can be defined and implemented. So, being the first drawing algorithm using ellipses, each option was thoroughly evaluated before one that effectively aids in achieving our objectives was chosen. The rationale for our design choices is clearly explained, and experimental and pragmatic evidence justifies our choices for each of the components. The type of diagrams that eulerAPE was aimed at generating is outlined later on. This justifies why analytic geometry methods were developed and used to compute the region areas of the three intersecting ellipses.

To evaluate the effectiveness of eulerAPE and ellipses in drawing accurate area-proportional Venn diagrams with smooth curves for most data with three sets (objective 2), we:

- evaluated the effectiveness of eulerAPE in drawing an accurate diagram when one is known to exist for the given area specification (Section 5.7.1);

- evaluated the effectiveness of eulerAPE in drawing an accurate diagram for any area specification for which an accurate diagram drawn with ellipses might not exist and compared these diagrams with those generated by a variant of eulerAPE that restricts the ellipses to circles (Section 5.7.2);

- identified characteristics of the area specifications for which an accurate diagram can or cannot be drawn using ellipses (Section 5.7.2);

- compared the effectiveness of eulerAPE and venneuler [Wilkinson, 2012] (considered to be the most effective circle-based drawing method) in generating accurate Venn diagrams for a set of random area specifications (Section 5.7.3);

- compared the quality of the diagrams generated by eulerAPE and various other drawing methods that use circles or polygons in depicting real data for an application area (Section 5.7.4).

Design features that could facilitate comprehension of such diagrams (objective 3) were suggested after looking into theories of perception and cognitive for visual design. We will now explain: the rationale for focusing on 3-Venn diagrams rather than Venn or Euler diagrams with any number of curves; the diagram types that eulerAPE aims at generating. These clarify our objectives for this chapter.

Why 3-Venn Diagrams and not n-Venn or n-Euler Diagrams

This work focuses on 3-Venn diagrams rather than n-Venn or n-Euler diagrams for various reasons.

Venn diagrams with three curves are the most commonly used area-proportional Euler diagrams (Section 3.7.1), but cannot be drawn accurately using circles (Section 3.7.6). Thus the need for a method that automatically draws accurate 3-Venn diagrams with smooth, distinguishable curves like circles.

An n-Venn diagram with n curves depicts all the possible intersections between n curves, while an n-Euler diagram depicts any intersection, containment or exclusion between n curves (Section 3.4). So there is one abstract description corresponding to a Venn diagram with n curves, but there could be various for an Euler diagram with the same n number of curves (Section 3.4.2). There could also be different ways how a diagram with a specific abstract description could be drawn (e.g., Figure 3.12). If

the diagram is area-proportional, any quantitative data could be assigned to the zones in the diagram (Section 3.7.2). Thus, if area-proportional Euler diagrams are handled, various diagrams, variables and quantitative data for each of the different possible abstract descriptions would have to be evaluated. Our algorithm eulerAPE is the first to use ellipses. Drawing accurate area-proportional 3-Venn diagrams with respect to an area specification using a specific shaped curve is a well-known, difficult problem (Section 3.7.6) and using ellipses makes the problem more complex [Chow and Ruskey, 2004]. Focusing on diagrams with one abstract description and a fixed number of curves (3; the most common) aid in **devising an effective algorithm that is the first to draw area-proportional diagrams with ellipses**.

Evaluating the effectiveness of ellipses in drawing area-proportional diagrams with one abstract description and a fixed number of curves **facilitates the identification of the area specification types for which an area-proportional diagram cannot be drawn with ellipses**. Example, there are area specifications for which an accurate area-proportional wellformed 3-Venn diagram cannot be drawn with convex curves (Section 3.7.5). Wellformedness (Section 3.5.1, Section 3.7.3) and curve convexity (Section 3.5.2) facilitate diagram understanding, so such properties should be met when possible. Thus focusing our evaluation on 3-Venn diagrams with ellipses aids in identifying such area specifications.

There are abstract descriptions for which a not area-proportional Euler diagram can only be drawn using non-simple curves (e.g., Figure 3.9B) **or irregular curves** (e.g., Figure 3.9D), **so ellipses cannot be used**. A few of the circle-based drawing methods still attempt to draw diagrams for such area specifications. However, the generated diagrams are far more misleading than ones with incorrect zone areas, as incorrect zones thus set relations are depicted. Example, Figure 3.28D by venneuler is a Venn diagram despite that an Euler diagram for abstract description $\{\varnothing, a, b, c, ab, abc\}$ was required. In fact, an Euler diagram for this abstract description can only be drawn using irregular curves, as Figure 3.9D.

For some abstract descriptions, a not area-proportional Euler diagram that is wellformed can only be drawn if the diagram has empty zones (Section 3.4.3), **so quantitative data cannot be depicted by the zone areas of such diagrams**. A non-wellformed diagram could be used, but the diagram would be difficult to comprehend particularly if it has disconnected zones (Section 3.7.3). Thus for some abstract descriptions, it is not possible to draw an area-proportional Euler diagram that effectively conveys the required set relations and quantitative data.

Rodgers et al. [2014] have devised a method that draws area-proportional Euler diagrams with up to three curves using circles when possible, convex polygons if circles cannot be used, and non-convex polygons if convex polygons cannot be used. The authors show that **for most abstract descriptions and area specifications, an accurate area-proportional 3-Euler diagram can only be drawn using non-smooth polygons that are not easily distinguishable as they overlap and make the diagram non-wellformed** (only 37.5% of the abstract descriptions corresponding to an Euler diagram with up to three curves can be drawn using circles for any area specification), **so ellipses cannot be used**.

Current drawing methods indicate the difficulty in devising measures that adequately identify zone area inaccuracies as well as missing or not required zones, to generate diagrams that accurately depict the required set cardinalities and relations. Example, as noted in Section 5.2.1, venneuler's *stress* could be relatively low (indicating that a diagram is good for the required area specification) even when the diagram has missing or unwanted zones, and thus incorrectly depicts the required set relations. Focusing on one abstract description (that of a Venn diagram; using shading a Venn depicts an Euler diagram) and a fixed number of curves (3; the more common) also **simplifies the development and thorough evaluation of all the components that make an effectiveness drawing algorithm that uses ellipses.**

For the above reasons, this work focused on 3-Venn diagrams. eulerAPE draws some Euler diagrams, but further investigations have be carried out in the future.

The Type of Diagrams eulerAPE *Aims at Generating*

Various aesthetic features that could affect the comprehension of Venn and Euler diagrams have not been studied empirically and no aesthetic metrics have been formalized (Section 3.5, Section 3.7.3). Chow and Ruskey [2004] argue that the curves in which a zone is located should be easily identified and the zone areas should be easily comparable to one another. To ensure that the curves are smooth, convex and distinguishable (Section 3.5.2), eulerAPE represents the curves as ellipses. Wellformedness facilitates diagram understanding (Section 3.5.1), particularly when the diagram is area-proportional (Section 3.7.3). So eulerAPE aims to draw diagrams that are wellformed.

Studies indicate that humans are biased to area judgement, yet it is still unclear how areas in such diagrams are perceived and what discrepancies in zone area are not noted (Section 3.7.3). Perceptual scaling was proposed for map symbols to help readers perceive their areas correctly, but studies show that no universal scaling metrics can be defined (Section 3.7.3). Such scaling measures were harshly criticised by Tufte who claimed that "graphical excellence begins with telling the truth about the data" [Tufte, 1983, p. 53] and if area is used to depict quantitative data, then the represented area "should be directly proportional to the numerical quantities represented" [Tufte, 1983, p. 56]. Tufte also defined the 'Lie Factor' formula to determine whether a visualization is representing the quantitative data accurately (Section 3.7.3). So for eulerAPE, a good, accurate diagram is one whose zone areas are all directly proportional to the quantities in the given area specification (Section 5.6.3).

However, this is only possible if the zone areas of the diagram are computed accurately. Numerical approximate methods introduce error [Antia, 2002; Isaacson, 1994] that could distort the visualization and its integrity, violating Tufte's [1983] graphical integrity principle. So, eulerAPE uses an analytic geometry algorithm to accurately compute the region areas of the three intersection ellipse, which, similar to numerical methods, provide results instantaneously (in 10 milliseconds) (Section 5.4).

The shape of the curves and zones could affect judgement, comparison and estimation of their area, but their effect has not been studied empirically as yet (Section 3.7.3). In some cases, it is not possible to draw a diagram that accurately depicts the required quantitative data and satisfies desired aesthetics, yet it is still unclear what a good compromise between zone area accuracy and aesthetics for diagram comprehension could be (Section 3.7.3). So currently the concept of accuracy is objective, but that of the aesthetics is still highly subjective as various empirical studies are yet to be carried out.

Hence, considering the importance for the "representation of numbers, as physically measured on the surface of the graphic itself" [Tufte, 1983, p. 56], eulerAPE aims at generating **wellformed diagrams with smooth, convex curves represented as ellipses and with zone areas that are accurately and directly proportional to the represented quantitative data**, thus conforming with Tufte's graphical integrity principle. Once area specifications for which an accurate diagram can be drawn are identified (whose zone areas are computed using analytic methods) and once a better understanding of diagrams' aesthetics is obtained, constraints on the zone area accuracy could be eased to satisfy important aesthetics based on their priority. Our approach is consistent with that of other methods, including the circle-based `venneuler` that aims to generate diagrams whose zone "areas are proportional to the size of subsets" [Wilkinson, 2012, p. 337] without taking into account any specific aesthetic considerations.

We will now discuss basic concepts and definitions, namely those related to ellipses (Section 5.3).

5.3 Basic Concepts and Definitions

The basic concepts and definitions in relation to area-proportional Venn diagrams are discussed in Section 3.7.2. A wellformed diagram is defined in Section 3.4.4. So this section focuses on ellipses.

5.3.1 Ellipses

An *ellipse* is a simple closed curve characterized by (Figure 5.5A): a centre, (γ_1, γ_2); two semi-axes, α and β, where $\alpha \geq \beta$; and an angle of rotation, θ, where $0 \leq \theta < 2\pi$. Since $\alpha \geq \beta$, often α and β are respectively referred to as the semi-major axis and semi-minor axis.

An ellipse is in *canonical form* if $(\gamma_1, \gamma_2) = (0, 0)$ and $\theta = 0$, so that in the Cartesian coordinate system, the ellipse is centred on the origin and the ellipse's semi-axes are along the *x*-axis and *y*-axis, as shown in Figure 5.5B. An ellipse with any (γ_1, γ_2) and θ is *a general ellipse*.
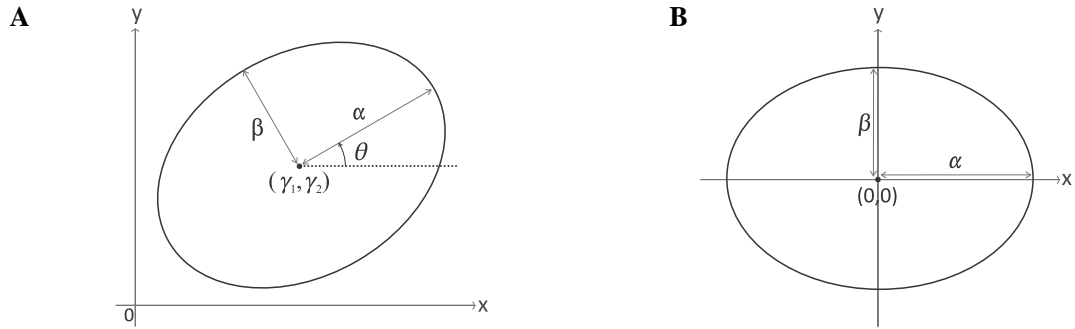
Figure 5.5: *An ellipse and its properties.* (A) An ellipse is characteristic by: a centre, (γ_1, γ_2); two semi-axes, α and β, *where* $\alpha \geq \beta$; an angle of rotation, θ. (B) An ellipse in canonical form has $(\gamma_1, \gamma_2) = 0$ and $\theta = 0$.

The *area of an ellipse* is

$$\pi\alpha\beta \tag{5.1}$$

The *curve of an ellipse*, using a *polar coordinate system* with pole (or origin) (γ_1, γ_2) and polar axis a ray from (γ_1, γ_2) passing through α, is defined as

$$\rho^2 = \frac{\alpha^2 \beta^2}{\beta^2 \cos^2 \varphi + \alpha^2 \sin^2 \varphi} \tag{5.2}$$

where φ is the polar angle between α and
a ray from (γ_1, γ_2) passing through a point on the ellipse

An *ellipse arc* ($\frown$) is a connected portion of the ellipse curve (e.g., $\overset{\frown}{MN}$ in Figure 5.6A).

An *ellipse sector* ($\triangledown$) is the space bounded by an ellipse arc and two line segments between the ellipse's centre and the arc's endpoints (e.g., $\triangledown MNO$ in Figure 5.6B). The *area of an ellipse sector* can be defined using the polar coordinates representation of the ellipse curve in Equation (5.2) as

$$\frac{1}{2}\int_{\varphi_1}^{\varphi_2} \rho^2\, d\varphi = \frac{\alpha^2 \beta^2}{2}\int_{\varphi_1}^{\varphi_2} \frac{d\varphi}{\beta^2 \cos^2\varphi + \alpha^2 \sin^2\varphi}$$

$$= \frac{\alpha\beta}{2}\tan^{-1}\left(\frac{\alpha}{\beta}\tan\varphi\right)\Bigg]_{\varphi_1}^{\varphi_2} \tag{5.3}$$

where φ_1 and φ_2 are the polar angles of the arc's endpoints from α with respect to (γ_1, γ_2)
and $0 < \varphi_2 - \varphi_1 \le 2\pi$

The value returned by Equation (5.3) is the area of the ellipse sector from φ_1 to φ_2 in an anticlockwise direction along the ellipse. So, to compute the area of $\diamond MNO$ in Figure 5.6B using Equation (5.3), φ_1 and φ_2 should respectively be the polar angles of M and N from α with respect to (γ_1, γ_2). In Figure 5.6B, α is along the x-axis and (γ_1, γ_2) is $(0,0)$ or O. If instead, φ_1 and φ_2 are respectively the polar angles of N and M from α with respect to O, the area computed by Equation (5.3) is the area of the ellipse minus the area of $\diamond MNO$.

An *ellipse segment* ($\frown$) is the space bounded by an ellipse arc and a chord that share the same endpoints (e.g., $\frown MNO$ in Figure 5.6C). As shown in Figure 5.6D, an ellipse sector ($\diamond MNO$), is composed of an ellipse segment ($\frown MNO$) with the same ellipse arc as the sector and a triangle ($\triangle MNO$). Thus, the *area of an ellipse segment* can be defined as

$$\text{Area of } \frown MNO = \text{Area of } \diamond MNO - \text{Area of } \triangle MNO \tag{5.4}$$

The area of $\diamond MNO$ in Equation (5.4) is computed using Equation (5.3) for the same arc endpoints as that of $\frown MNO$.
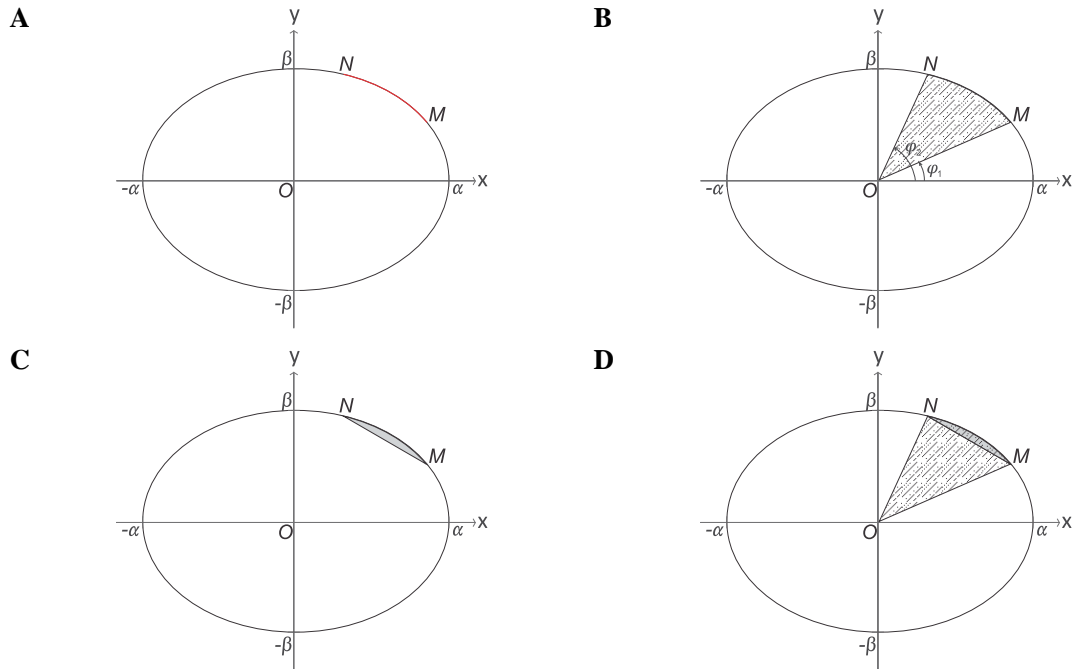
Figure 5.6: *An ellipse arc, an ellipse sector, an ellipse segment, and an ellipse sector with an ellipse segment.* (a) $\widehat{MN}$ (in red) is an ellipse arc between points M and N on the ellipse curve. (b) $\diamond MNO$ (dashed) is an ellipse sector defined by $\widehat{MN}$ and line segments OM and ON. φ_1 and φ_2 are respectively the polar angles of M and N from α with respect to O. (c) $\frown MNO$ (in grey) is an ellipse segment defined by $\widehat{MN}$ and chord MN. (d) $\diamond MNO$ is made up of $\frown MNO$ and $\triangle MNO$, so the area of $\frown MNO$ = area of $\diamond MNO$ − area of $\triangle MNO$.

The polar coordinates representation of an ellipse curve in Equation (5.2), assumes that (γ_1, γ_2) is at the pole and α is along the polar axis of the coordinate system. So Equation (5.2) cannot represent a general ellipse.

The *curve of a general ellipse* can be defined *parametrically* as

$$x(t) = \gamma_1 + \alpha \cos\theta \cos t - \beta \sin\theta \sin t$$
$$y(t) = \gamma_2 + \alpha \sin\theta \cos t + \beta \cos\theta \sin t$$

(5.5)

where t is the parameter and $0 \le t \le 2\pi$ and
$(x(t), y(t))$ are the Cartesian coordinates of a point on the ellipse curve

In Equation (5.5), t is the angular parameter (or eccentric angle or eccentric anomaly) that determines the position of a particle moving along the ellipse curve, so that every value of t determines the Cartesian coordinates $(x(t), y(t))$ of a point on the ellipse curve. As shown in Figure 5.7, given a point (x, y) on the ellipse curve, the corresponding value of t is determined by drawing a line perpendicular to α that passes through (x, y) and intersects with the auxiliary circle (i.e., a circle with radius α and centre (γ_1, γ_2) that is the circumscribed circle of the ellipse; in red in Figure 5.7) at a point P, and by then computing the anticlockwise angle from α to the line passing through (γ_1, γ_2) and P with respect to (γ_1, γ_2). The latter angle is the value of t. Since in Figure 5.7, $\theta = 0$ and point (x, y) is on the top right quarter of the ellipse curve, the value of t for (x, y) can be computed using any one of the following

$$t = \cos^{-1}\left(\frac{x - \gamma_1}{\alpha}\right) \quad \text{or} \quad t = \sin^{-1}\left(\frac{y - \gamma_2}{\beta}\right)$$

(5.6)

If $\theta = 0$ and the point (x, y) is on any other quarter of the ellipse curve, the above equations would need to be adapted. For instance, if the point (x, y) is on the upper left quarter of the ellipse curve, t is π minus the value of one of the above equations of t. If $\theta \ne 0$, the point (x, y) has to be rotated by $-\theta$ about (γ_1, γ_2) before the above equations of t or any adaptations of them can be used.
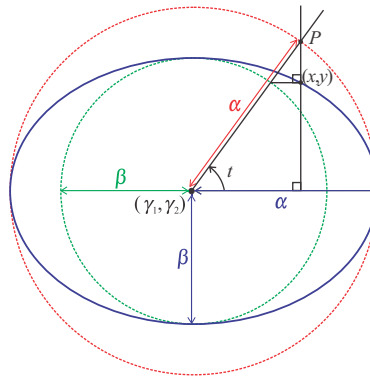
Figure 5.7: *Computing the value of parameter t of the parametric representation of the curve of a general ellipse, given the Cartesian coordinates (x,y) of a point on the ellipse curve.* The red circle with radius α and centre $(\gamma1, \gamma2)$ is the auxiliary circle that is the circumscribed circle of the ellipse (in blue). The green circle with radius β and centre $(\gamma1, \gamma2)$ is the minor auxiliary circle that is the inscribed circle of the ellipse (in blue). Both circles and the ellipse are concentric. The value of t for point (x, y) is the anticlockwise angle from α to the line passing through $(\gamma1, \gamma2)$ and P with respect to $(\gamma1, \gamma2)$. P is the point where the line perpendicular to α passing through (x, y) intersects the auxiliary circle (in red).

Alternatively, the *curve of a general ellipse* can be defined by the set of points (x,y) on the Cartesian plane that satisfy the following *implicit polynomial equation*:

$$\frac{\left((x-\gamma_1)\cos\theta+(y-\gamma_2)\sin\theta\right)^2}{\alpha^2}+\frac{\left((y-\gamma_2)\cos\theta-(x-\gamma_1)\sin\theta\right)^2}{\beta^2}=1 \tag{5.7}$$

This is rather complex to handle compared to the parametric representation in Equation (5.5) and the implicit polynomial equation that should be satisfied by the set of points (x,y) that define the curve of an ellipse in canonical form, that is

$$\frac{x^2}{\alpha^2}+\frac{y^2}{\beta^2}=1 \tag{5.8}$$

These basic concepts and definitions will be referred to in this chapter, particularly the next section where we discuss our geometry analytic methods to compute the zones areas of our diagrams drawn with ellipses. Further details about ellipses are available in geometry textbooks (e.g., [Eberly, 2011a; Farin et al., 2002; Page, 2009; Schneider and Eberly, 2002; Stewart, 2009]).

5.4 Analytic Geometry Methods for Computing the Region Areas of Three General Overlapping Ellipses

The region areas of two or three intersecting circles can be computed using analytic methods [Chow and Rodgers, 2005; Chow and Ruskey, 2004]. For more than three circles, numerical approximations [Antia, 2002; Schneider and Eberly, 2002] are often used, as analytic solutions for these cases are more complex. Example, `venneuler` [Wilkinson, 2012] uses a numerical quadrature method whereby each circle is drawn on a 200×200 byte plane and each 'pixel' on each plane is set to 1 if it is in the circle or 0 if not. The area of each region in the diagram is the sum of the value of all the pixels on all the planes that are located in that region. For diagrams with more than three circles, numerical methods can be computationally less expensive than analytic ones. However, approximation errors could be introduced and unless the required function is evaluated at an excessive number of intervals, the estimated value of the function could be unreliable [Antia, 2002; Isaacson, 1994]. When the curves are represented as polygons, the area of the overlaps is often computed by applying the Gaussian integration theorem [Kestler et al., 2005, 2008].

None of the drawing methods before `eulerAPE` used ellipses, possibly due to difficulties in calculating the area of the regions of overlapping ellipses [Chow and Ruskey, 2004]. We discuss possible methods that could be used and the method we devised for `eulerAPE`.

5.4.1 Possible Methods

Ellipses have more degrees of freedom than circles and so, using **an analytic method** to compute the region areas of overlapping ellipses might seem complex and too computationally expensive. However, an analytic method [Hughes and Chraibi, 2011] has been successfully used and efficient enough to compute the area of two overlapping ellipses for simulations of dynamic systems, (e.g., an orbiting satellite with a solar calibrator [Kent et al., 2009]; a force-based model for pedestrian dynamics [Chraibi et al., 2010; Chraibi, 2012]) and in image feature matching and extraction [Barandiaran et al., 2013a, b].

If eulerAPE had to use **a numerical approximation** such as venneuler's [Wilkinson, 2012] numerical quadrature method, errors would be introduced in the calculation of the region areas of the intersecting ellipses. Example, in Figure 5.8A and Figure 5.8B, the same diagram is drawn on two same 200×200 grids, with a difference in the placement of the diagrams on the grids. The magnified views of zones *ab*, *ac*, *bc* and *abc* indicate that according to Figure 5.8A, there are 0 grid squares or pixels that are only in zone *bc,* while according to Figure 5.8B, there are 3. Similar differences can be noted for the other zones, due to a difference in the positioning of the diagrams on the grids. Though using a larger grid could reduce inaccuracies, such issues will still be evident. So with this method, the zone areas are not computed accurately and small zones (e.g., zone *bc* in Figure 5.8) could be considered missing even though they are depicted. This could thus impede the optimization process from taking a path that could lead to an improved solution and impede the method from handling diagrams with small zones. The zone area inaccuracies would violate Tufte's primary graphical integrity principle [Tufte, 1983, p. 56], thus our objectives (Section 5.2.3) in generating diagrams whose zone areas are directly proportional to the required quantitative data and in identifying area specifications for which an area-proportional diagram can be drawn accurately using ellipses. So numerical methods are not appropriate for eulerAPE.
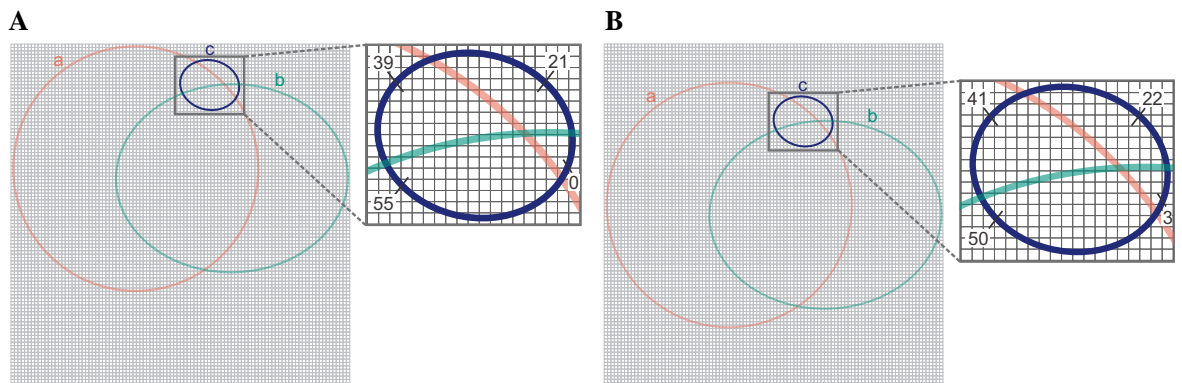
Figure 5.8: *Computing the region areas of intersecting ellipses using a numerical quadrature method.* The diagram and the 200×200 grid is the same for both A and B. The only difference is the positioning of the diagram on the grid, resulting in a different number of grid squares that are only inside a specific zone.

Also, to check if the pixel is in the circle, venneuler's numerical quadrature method [Wilkinson, 2012] takes the coordinates of the pixel's location as if it was a point on the Cartesian plane, and uses them and the implicit polynomial equation of the circle to verify if the distance between the circle's centre and the pixel is less than the circle's radius. If the latter is true, the pixel is in the circle. Checking if such a pixel is in an ellipse rather than a circle is more complex and conducting this check for all the pixels on all the planes would make the method more computationally expensive than it is for circles. This is so, as different from circles, ellipses have two semi-axes and an angle of rotation, and the implicit polynomial equation of the curve of a general ellipse (Equation (5.7)) is more complex. If instead, the implicit polynomial equation of the curve of an ellipse in canonical form (Equation (5.8)) is used, both the ellipse and the pixel would have to be translated and rotated before this simpler equation is used. An alternative would be to have a secant line passing through the ellipse's centre and the pixel to be checked, then compute the intersection points of this secant line with the ellipse, identifying the intersection points that has the same polar angle from the semi-axes α with respect to ellipse's centre as that of the pixel. If the distance between the ellipse's centre and this intersection point is greater than the distance between the ellipse's centre and the pixel, then the pixel is in the ellipse. However, all of these methods to check if a pixel is in an ellipse are more computationally expensive than those for circles.

Alternatively, the ellipses could be represented using regular convex polygons and **a standard polygon intersection algorithm** could be used to compute the area of the regions. However, once again, the zone areas would not be computed accurately, small zones could be considered missing, and unless the polygons has numerous vertices and short edges, the computed zone areas would be unreliable approximations of the zones areas of intersecting ellipses, as the curves would lack the good continuity of ellipses. Polygon intersection algorithms for polygons with numerous vertices are also more computationally expensive than the numerical quadrature method [Schneider and Eberly, 2002].

Hence, to achieve our objectives (Section 5.2.3), namely that of accuracy, **we opted for an analytic method** that would compute the accurate area of the regions of the intersecting ellipses. As noted in Section 5.4.4, the analytic method used by eulerAPE provides results in 10 milliseconds (on an Intel Core i7 CPU @2GHz with 8GB RAM, OS X 10.8.4, Java Platform 1.6.0_51), which is 10 times less than the 0.1 second limit for an instantaneous response [Card et al., 1991; Miller, 1968] and which is relatively similar to the 1 millisecond (on a MacBook Pro @2.5Ghz with 2GB RAM, Java Platform 1.5) that venneuler's [Wilkinson, 2012] numerical quadrature method takes to provide the area of the regions of intersecting circles. Thus, an analytic method can still be efficient and yet ensure zone area accuracy.

Though venneuler's [Wilkinson, 2012] numerical quadrature method is relatively fast, the computed zone areas are not accurate and so, the analytical gradient of its steepest descent optimization method, in search for a diagram with respect to an area specification, cannot be computed. Instead, an approximation of the gradient is used. This could be expensive to compute as various iterations involving repeated evaluation of venneuler's statistical loss function, *stress*, for each curve are required to obtain a good approximation of the true gradient. So, after all, a similar amount of time is required to generate a diagram whose zone areas are computed using either a numerical method or an analytic method. This is demonstrated in our comparative analysis of eulerAPE and venneuler in Section 5.7.2.

5.4.2 Current Analytic Methods

Two analytic methods are available to compute the region areas of the overlapping ellipses: one by Eberly [2010] and another by Hughes and Chraibi [2011] (both as non-refereed articles). Both methods are restricted to two ellipses. Eberly's method is further restricted to ellipses in canonical form, while that by Hughes and Chraibi can handle any two general ellipses with any centre and angle of rotation. Both methods compute the area of the overlapping region by first obtaining the area of the two ellipse segments compromising the region (as shown later in Figure 5.9B). The area of each ellipse segment is obtained by computing the area of an ellipse sector and then subtracting the area of a triangle, as shown earlier in Figure 5.6D. The derivation of the area of a sector of an ellipse in canonical form is obtained using integral calculus. The representation of the ellipse curve is defined in polar coordinates by Eberly and parametrically by Hughes and Chraibi. To handle general ellipses, Hughes and Chraibi first translate and rotate the general ellipses so they are transformed into canonical form, and then compute the area of the required ellipse segment using the same equation as that of ellipses in canonical form.

However, no analytic methods that compute the region areas of three general intersecting ellipses have been proposed. So we developed, implemented and evaluated two different and efficient analytic methods, one of which is used by eulerAPE. We explain our methods in the next section. It should be noted that only Eberly's method was available at the time our method was developed. Hughes and Chraibi's method was available months after our analytic methods were developed and implemented.

5.4.3 Our Analytic Methods

We devised **two general analytic methods** for three ellipses, both of which can handle ellipses that are not necessarily in canonical form. These include:

- M1. Decomposition into ellipse segments;
- M2. Using integral calculus.

These methods differ in the way they compute the area of the overlapping region between two ellipses and that of an ellipse segment. Similar to Eberly's [2010] and Hughes and Chraibi's [2011] methods, M1 decomposes the region of interest into ellipse segments and uses an equation of the segment area of an ellipse in canonical form. M2 uses integral calculus to directly derive the equation of the required enclosed region area without any transformations. M1 and M2 are discussed later.

The **general algorithm** for computing the region areas of three intersecting ellipses representing a Venn diagram using either M1 or M2 is as follows:

Algorithm 5.1: *ComputeRegionAreasOfThreeIntersectingEllipses* (*d*)

Input: A Venn diagram *d* with three ellipses
Output: *regionAreas*, a set of areas one for every region in *d*

1. **for** each ellipse *e* in *d* **do**
2.   *ellipseAreas*[*e*] ← area of *e* by Equation (5.1)
3. **end for**
4. **for** each pair of ellipses e_1 and e_2 in *d* **do**
5.   *points*[e_1, e_2] ← intersection points of e_1 and e_2 by Hill's [1995] method
6.   *overlapAreas*[e_1, e_2] ← area of the overlapping region between e_1 and e_2
                                 by M1 or M2 and *points*[e_1, e_2]
7.   *interiorPoints*[e_1, e_2] ← the intersection point in *points*[e_1, e_2] that
                                 is inside the third ellipse of *d*
8. **end for**
9. Decompose the region in all three ellipses of *d* into ellipse segments es_1, es_2, es_3 and triangle *t* (as in Figure 5.9D) using *interiorPoints*, which defines the arc endpoints of es_1, es_2, es_3 and the vertices of *t*
10. *regionAreas*[e_1, e_2, e_3] ← $\sum$ areas of es_1, es_2, es_3 by Equation (5.4) and *t*
11. **for** each pair of ellipses e_1 and e_2 in *d* **do**
12.   *regionAreas*[e_1, e_2] ← *overlapAreas*[e_1, e_2] − *regionAreas*[e_1, e_2, e_3],
                            where e_3 is the other ellipse in *d*
13. **end for**
14. **for** each ellipse *e* in *d* **do**
15.   *otherAreas* ← $\sum$ *regionAreas*[*e*, e_1], *regionAreas*[*e*, e_2], *regionAreas*[*e*, e_1, e_2],
                    where e_1 and e_2 are the two other ellipses in *d*
16.   *regionAreas*[*e*] ← *ellipseAreas*[*e*] − *otherAreas*
17. **end for**
18. **return** *regionAreas*

If the diagram represented by the intersecting ellipses is wellformed, all of its zones are connected and so each of the mentioned regions in Algorithm 5.1 correspond to precisely one zone. The area of these regions is thus the area of the corresponding zone.

Algorithm 5.1 has been implemented for both M1 and M2 and for any possible representation of a wellformed 3-Venn diagram drawn using ellipses that are not necessarily in canonical form. M1, M2 and the chosen method to compute the ellipses' intersection points are discussed in the next sections.

If the general intersecting **ellipses do not form a wellformed 3-Venn diagram**, an extended version of Algorithm 5.1 would have to be implemented, so that the various ways how the ellipses can intersect are handled (there can be from zero up to four intersection points between two ellipses [Eberly, 2010]). The method we have chosen to compute the intersection points [Hill, 1995] returns all the intersection points (i.e., zero up to four) between any two general ellipses. So the algorithm can easily be extended by: (i) identifying the way each pair of ellipses intersect from the number of intersection points between the two ellipses; (ii) decomposing the relevant regions into ellipse segments and basic geometry shapes like triangles or rectangles when necessary; (iii) using either M1 or M2 to find the area of the overlapping region between two ellipses and the area of ellipse segments; (iv) using Equation (5.1) to compute the area of the ellipses; (v) using basic algebra to add and subtract areas wherever necessary to calculate the region areas. In such cases, a zone in the diagram could be made up of multiple regions and so, its area would be the sum of the area of all of these regions. At present, eulerAPE draws 3-Venn diagrams that are wellformed and thus, such an extension is currently not required.
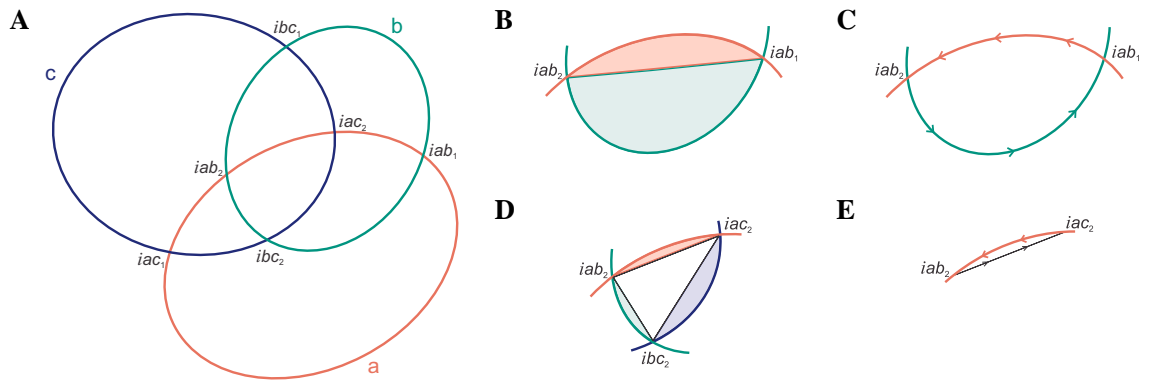
Figure 5.9: *Computing the area of the overlapping region between two ellipses and the area of the region that is located in exactly the three ellipses using M1 and M2.* (A) Three general intersecting ellipses representing a wellformed 3-Venn diagrams with curves *a*, *b* and *c*. The intersection points of every ellipse pair are labelled. Example, iab_1 and iab_2 are the intersection points of *a* and *b*. (B) The overlapping region between *a* and *b* in A decomposed into two ellipse segments. M1 computes the area of such a region by adding the area of the two ellipse segments. (C) The overlapping region between *a* and *b* in A. This region is the same as that in B, but the region is not decomposed into ellipse segments. M2 computes the area of such a region by adding the area under the ellipse curve *a* from iab_1 to iab_2 and the area under the ellipse curve *b* from iab_2 and iab_1, where each area is defined by a definite integral. (D) The region in A that is located in exactly the three ellipses. M1 and M2 decompose the region into ellipse segments (a segment of ellipse *a* with arc endpoints iac_2 and iab_2, a segment of ellipse *b* with arc endpoints iab_2 and ibc_2, a segment of ellipse *c* with arc endpoints ibc_2 and iac_2) and a triangle (with vertices iac_2, iab_2 and ibc_2), and add up their areas. (E) The segment of ellipse *a* with arc endpoints iac_2 and iab_2 in D. M2 computes the area of such a segment by adding the area under the ellipse curve *a* from iac_2 to iab_2 and the area under the secant line passing through the arc endpoints from iab_2 and iac_2.

M1: Decomposition Into Ellipse Segments

To compute the area of the overlapping region of two general ellipses, M1 decomposes the region into two ellipse segments, as in Figure 5.9B. As explained earlier in Algorithm 5.1 and Figure 5.9D, the region in exactly the three ellipses is similarly decomposed into ellipse segments and a triangle.

Equations defining the area of a segment of an ellipse in canonical form can be easily derived and could be less complex than ones for a general ellipse. Such equations are already available. Example, Eberly [2010] and Hughes and Chraibi [2011] provide two different equations based on the area of a sector of an ellipse in canonical form. Eberly's equation uses the polar coordinates representation of the ellipse curve, while that of Hughes and Chraibi uses the parametric representation of the ellipse curve.

M1 uses Equation (5.4) to compute the area of an ellipse segment which, similar to Eberly's equation, uses the polar coordinates representation of the ellipse curve in Equation (5.2). As shown in Figure 5.9B and Figure 5.9D, the arc endpoints of an ellipse segment are two of the intersection points of the three overlapping ellipses, which define the arc in an anticlockwise direction along the ellipse (the importance of this direction is explained in Section 5.3.1 in relation to Equation (5.3)). The arc endpoints and the ellipse are rotated by an angle of $-\theta$ about (γ_1, γ_2) and translated to $(0,0)$, so that the transformed ellipse is in canonical form. Rotation and translation are affine transformations and so, the ellipse's properties, its area and the area of the required ellipse segment are preserved. The polar angle of each transformed endpoint from α with respect to $(0,0)$ is then computed, and together with the transformed ellipse, these polar angles are used to compute the area of the required ellipse segment using Equation (5.4).

M2: Using Integral Calculus

M2 derives an equation of the area of the required enclosed region from definite integrals that compute the area under an ellipse curve (or line) between two given points. The equation of the curve of a general ellipse is used and so, M2 does not require any of the geometric transformations used in M1 to get the ellipse in canonical form.

Let e_1 and e_2 be two ellipses that intersect at points i_1 and i_2, such that the overlapping region is enclosed by an ellipse arc from i_1 and i_2 in an anticlockwise direction along e_1 and an ellipse arc from i_2 and i_1 in an anticlockwise direction along e_2. In M2, the area of the overlapping region between e_1 and e_2 is defined as

$$\text{Area under curve } e_1 \text{ from } i_1 \text{ to } i_2 \;+\; \text{Area under curve } e_2 \text{ from } i_2 \text{ to } i_1 \tag{5.9}$$

This is illustrated in Figure 5.9C for ellipses a and b in Figure 5.9A, where a and b are respectively e_1 and e_2 in Equation (5.9), and iab_1 and iab_2 are respectively i_1 and i_2 in Equation (5.9).

To compute the area of the region located in exactly the three ellipses, M2 decomposes the region into ellipse segments and a triangle, as in M1 and as shown in Figure 5.9D, where the arc endpoints of the ellipse segments and the vertices of the triangle are defined by the intersection points of the three overlapping ellipses. Given a segment of an ellipse e with an arc from i_1 and i_2 in an anticlockwise direction along e and a secant line l intersecting e at i_1 and i_2, M2 defines the area of the ellipse segment as

$$\text{Area under curve } e \text{ from } i_1 \text{ to } i_2 \;+\; \text{Area under line } l \text{ from } i_2 \text{ to } i_1 \tag{5.10}$$

This is illustrated in Figure 5.9E for the segment of ellipse a between iac_2 and iab_2 that is located in exactly the three ellipses of Figure 5.9A, where a is e in Equation (5.9) and iac_2 and iab_2 are respectively i_1 and i_2 in Equation (5.9).

The area under a curve between two given points is defined by a definite integral. The curve of a general ellipse can be defined in different ways (Section 5.3.1). However, it is not always so straightforward to handle some representations. For instance, the implicit polynomial in Equation (5.7) has to be converted to an explicit polynomial before it can be used to find the area under an ellipse curve. The parametric representation in Equation (5.5) can be used as is and determining the antiderivative of this representation is straightforward. So M2 uses the parametric representation of the curve of the general ellipse in Equation (5.5), that is $x(t)$ and $y(t)$, and defines the area under the curve of a general ellipse e from (x_1,y_1) to (x_2,y_2) on e as follows:

If $|x_1-x_2|>|y_1-y_2|$

   If $y = F(x)$ is the explicit definition of e as a function of x and $x(t_1) = x_1$ and $x(t_2) = x_2$, the area under the curve e from (x_1,y_1) to (x_2,y_2) is

$$
\begin{aligned}
\int_{x_1}^{x_2} F(x)\,dx = \int_{t_1}^{t_2} F(x(t))x'(t)\,dt \ \ &= \int_{t_1}^{t_2} y(t)x'(t)\,dt \\
&= \int_{t_1}^{t_2}\big(\gamma_2 + \alpha\sin\theta\cos t + \beta\cos\theta\sin t\big)\big(-\alpha\cos\theta\sin t - \beta\sin\theta\cos t\big)\,dt \\
&= K_1\sin 2t + K_2\sin t + K_3\cos 2t + K_4\cos t + K_5 t\big]_{t_1}^{t_2}
\end{aligned}
$$

(5.11)

$$
where \quad K_1 = \frac{\alpha\beta\cos 2\theta}{4}, \quad K_2 = -\gamma_2\beta\sin\theta, \quad K_3 = \frac{\big(\alpha^2+\beta^2\big)\sin 2\theta}{8},
$$
$$
K_4 = \gamma_2\alpha\cos\theta, \quad K_5 = -\frac{\alpha\beta}{2}
$$

Else (i.e., $|x_1-x_2|\le|y_1-y_2|$)

   If $x = F(y)$ is the explicit definition of e as a function of y and $y(t_1) = y_1$ and $y(t_2) = y_2$, the area under the curve e from (x_1,y_1) to (x_2,y_2) is

$$
\begin{aligned}
\int_{y_1}^{y_2} F(y)\,dy = \int_{t_1}^{t_2} F(y(t))y'(t)\,dt \ \ &= \int_{t_1}^{t_2} x(t)y'(t)\,dt \\
&= \int_{t_1}^{t_2}\big(\gamma_1 + \alpha\cos\theta\cos t - \beta\sin\theta\sin t\big)\big(-\alpha\sin\theta\sin t + \beta\cos\theta\cos t\big)\,dt \\
&= K_1\sin 2t + K_2\sin t + K_3\cos 2t + K_4\cos t + K_5 t\big]_{t_1}^{t_2}
\end{aligned}
$$

(5.12)

$$
where \quad K_1 = \frac{\alpha\beta\cos 2\theta}{4}, \quad K_2 = \gamma_1\beta\cos\theta, \quad K_3 = \frac{\big(\alpha^2+\beta^2\big)\sin 2\theta}{8},
$$
$$
K_4 = \gamma_1\alpha\sin\theta, \quad K_5 = \frac{\alpha\beta}{2}
$$

The value of t_1 and t_2 corresponding to (x_1,y_1) and (x_2,y_2) respectively is computed as discussed in Section 5.3.1 in relation to Equation (5.5) and Equation (5.6).

Similarly, the area under a line l from (x_1,y_1) to (x_2,y_2) on l where m and c are respectively the gradient and y-intercept of l is defined as:

If $|x_1-x_2|>|y_1-y_2|$

   If $y = mx+c$ explicitly defines l as a function of x, the area under l from (x_1,y_1) to (x_2,y_2) is

$$
\int_{x_1}^{x_2} (mx+c)\,dx = \frac{mx^2}{2} + cx \Bigg]_{x_1}^{x_2}
$$

(5.13)

Else (i.e., $\left| x_1 - x_2 \right| \le \left| y_1 - y_2 \right|$)

If $x = \dfrac{y - c}{m}$ explicitly defines l as a function of y, the area under l from (x_1, y_1) to (x_2, y_2) is

$$\frac{1}{m} \int_{y_1}^{y_2} (y - c)\, dy = \frac{1}{m} \left(\frac{y^2}{2} - cy \right) \Bigg]_{y_1}^{y_2} \qquad (5.14)$$

However, to compute the region area of the overlapping ellipses, the intersection points of the ellipses are required. We now discuss our chosen method to compute the intersection points.

Computing the Intersection Points of the Ellipses

All the intersection points of the ellipses making up a diagram can be obtained by computing the intersection points of every ellipse pair.

An ellipse is a curve and so, the various methods of computing the intersection points of two curves [Böhm et al., 1984; Farin et al., 2002; Manocha and Demmel, 1994; Schneider and Eberly, 2002] can be adapted for the intersection points of two ellipses. Numerical methods [Sederberg and Parry, 1986], such as Bézier and internal subdivision [Yap, 2006], Bézier clipping [Sederberg and Nishita, 1990] and the Newton-Raphson method, could also be used. However, as explained in Section 5.4.1, numerical methods are not appropriate for eulerAPE as zone area accuracy is important to conform with Tufte's primary graphical integrity principle [Tufte, 1983, p. 56] and to meet our objectives (Section 5.2.3).

The most common analytic methods include: (i) the resultant-based method using for instance Bezout's resultant [Eberly, 2010, 2011b]; (ii) the Gröbner basis method [Farin et al., 2002]; (iii) the matrix-based method [Hill, 1995]. These methods have been used for various areas (e.g., the resultant-based method [Hughes and Chraibi, 2011; Maitre and Nguyen, 2004]; the Gröbner basis method [Pernici, 2005]; the matrix-based method [Alfano and Greer, 2003]).

Methods (i) and (ii) are more complex than (iii) as the roots of a quartic polynomial have to be solved using Ferrari's solution (e.g., [Herbison-Evans, 1995; Korn and Korn, 1968]) or other methods (e.g., [Herbison-Evans, 1995]). Example, Hughes and Chraibi's [2011] analytic method, for the region areas of two intersecting ellipses, uses a numerical implementation of Ferrari's solution [Nonweiler, 1968] to find the roots of the quartic polynomial. However, we only require the real roots and so, finding all the roots of a quartic polynomial is a waste of computational resources [Herbison-Evans, 1995].

Using the matrix representation of conic sections and homogeneous transformation matrices, method (iii) can compute the intersection points of the two ellipses without referring to any quartic polynomials. This method has now been extended to efficiently identify whether two solid ellipsoid intersect [Alfano and Greer, 2003]. So eulerAPE uses this method [Hill, 1995] to compute the ellipses' intersection points.

5.4.4 Evaluation of Our Methods

M1, M2 and an approximate method were used to compute the zone areas of 8000 wellformed 3-Venn diagrams with ellipses whose properties were randomly generated. Some of these randomly generated diagrams had very small region areas that were barely visible. The computed areas were then compared. The approximate method represented the ellipses as regular convex polygons and used a standard polygon intersection algorithm to find the area of the regions and to compute the intersection points.

The same areas were obtained by our analytic methods, M1 and M2 (with an occasional insignificant difference of less than 10^{-4}). The average percentage error of the areas provided by the approximate method with respect to any one of our analytic methods was 1.04%. So the areas provided by M1 and M2 are similar to the approximate areas computed with a standard algorithm, and considering that the areas of M1 and M2 are the same, then these areas are after all correct.

M1 and M2 computed the areas in around 10 milliseconds on an Intel Core i7 CPU @2GHz with 8GB RAM, OS X 10.8.4 and Java Platform 1.6.0_51. This response time is 10 times less than the 0.1 second limit for an instantaneous response [Card et al., 1991; Miller, 1968], and similar to that of numerical methods (e.g., `venneuler`'s [Wilkinson, 2012] response time is 1 millisecond using a numerical quadrature method on a MacBook Pro @2.5Ghz with 2GB RAM and Java Platform 1.5). So our analytic methods can also be used for any other application, including simulations of dynamic systems, where both accuracy and efficiency is important.

5.4.5 The Method Employed By euler*APE*

As discussed in the previous section, there are no significant differences in terms of accuracy and efficiency between M1 and M2, so any one could be employed by euler*APE*. However, at present euler*APE* uses M2 so that the area of the regions is computed without the need of any geometric transformations.

We now explain how we generated libraries of random area specifications to evaluate our design choices for euler*APE* in Section 5.6 and euler*APE*'s effectiveness in achieving our objectives in Section 5.7.

5.5 Generating Different Libraries of Random Area Specifications

The various design options for euler*APE* were thoroughly evaluated (Section 5.6), before the most effective choices to meet our objectives were adopted. So experimental evidence supports our design rationale. Once euler*APE* was implemented, the effectiveness of euler*APE* and ellipses in drawing accurate area-proportional 3-Venn diagrams was evaluated and compared with other methods (Section 5.7).

Random area specifications for the above experiments were obtained from two types of libraries, DIAGLIB and AREASPECLIB, that were generated at both the design phase and the evaluation phase of euler*APE* (to avoid using the same library for the two phases). So in all we had four libraries of two types:

- DIAGLIB_*design* and AREASPECLIB_*design*, for the design phase;
- DIAGLIB_*eval* and AREASPECLIB_*eval*, for the evaluation phase.

From each library, 10,000 area specifications each corresponding to the abstract description of a 3-Venn diagram with curves labelled *a*, *b* and *c* were obtained. A Venn diagram has one abstract description, which in this case is $\{\varnothing, a, b, c, ab, ac, bc, abc\}$. So each area specification, ω, was of the form

$$\omega = \left\{(a, k_1), (b, k_2), (c, k_3), (ab, k_4), (ac, k_5), (bc, k_6), (abc, k_7)\right\}$$

$$where \ \ \left\{k_1, k_2, k_3, k_4, k_5, k_6, k_7\right\} \subset \mathbb{R}_{>0}$$

Conducting our experiments on a large sample of 10,000 area specifications at both the design phase and the evaluation phase was important to ensure that the sample is representative of the population of area specifications we were studying and to conform with the 'Law of Large Numbers', which states that the larger the sample, the more reliable the obtained findings, as its mean result would be closer to the expected value than that of a smaller sample [Bernoulli, 1713; Sedlmeier and Gigerenzer, 1997].

We will now explain how the libraries of type DIAGLIB and AREASPECLIB were generated and how the area specifications were obtained from each one, including details on the type of area specifications each represented and how they were used. Later, we discuss the characteristics of the area specifications in each of the four generated libraries, namely DIAGLIB_*design*, AREASPECLIB_*design,* DIAGLIB_*eval* and AREASPECLIB_*eval*.

5.5.1 Two Types of Libraries

DIAGLIB

A library of type DIAGLIB provides a sample of the population of area specifications for which an accurate wellformed area-proportional 3-Venn diagram can be drawn with ellipses. Such area specifications can be obtained by accurately computing the zone areas of randomly generated diagrams. The zone areas of the generated diagrams can be computed using our analytic methods in Section 5.4.3.

Hence, a library of this type consists of 10,000 random wellformed Venn diagrams drawn with three ellipses labelled *a*, *b* and *c*. For each ellipse *e*, random real positive numbers taken from a uniform distribution are used to define the properties of *e*, namely:

- centre (γ_1,γ_2), where (γ_1,γ_2) and the entire curve of *e* are in eulerAPE's diagram display panel;
- semi-axes α and β, where α and β are in the interval [5, 180] and $\alpha \geq \beta$,
  (180 pixels is a quarter of the width of eulerAPE's diagram display panel);
- angle of rotation θ, where θ is in the interval [0, π).

Though the restrictions on the value of (γ_1,γ_2) and α and β are practical, they are not indispensable as the diagrams can be translated and scaled to fit in eulerAPE's diagram display panel. Similarly, θ could be in the interval [0, 2π). However, we restricted θ to [0, π), as due to symmetry, an ellipse with an angle of rotation *x* in [0, π) is the same as an ellipse with an angle of rotation *x*+π, given that the two ellipses have the same centre and semi-axes. Three ellipses with randomly generated properties make up a diagram. If it is a wellformed Venn diagram, the diagram is included in the library, else it is discarded.

Generated diagrams with zone areas less than 1 pixel are not included in the library. Such areas are not depicted on a display device and so, an Euler diagram rather than a Venn diagram is displayed. Similarly, generated diagrams that would be displayed (on a display device) with disconnected zones or brushing points are not included in the library. To identify these diagrams, the ellipses are first represented as regular convex polygons (constructed by computing the value of the ellipse curve for every 0.5 interval in *x*) and then, polygons representing the zones are obtained. If a zone is made up of zero or more than one polygon with area greater than 1 pixel, the diagram is discarded.

Some of the generated diagrams, such as those in Figure 5.10, might not be considered aesthetically desirable, as the shape of some of their zones could hinder the estimation of their area and comparisons with other zone areas (Section 3.7.3). For instance, zones *b* and *bc* in Figure 5.10A have a very irregular shape, while zones *a* and *abc* in Figure 5.10B have a very elongated shape. However, this is still highly subjective as such aesthetic characteristics have not been studied empirically (Section 3.7.3). Since such diagrams are wellformed Venn diagrams, they are still included in the library as the area specification they are representing is one for which a wellformed accurate area-proportional 3-Venn diagram can be drawn with ellipses.
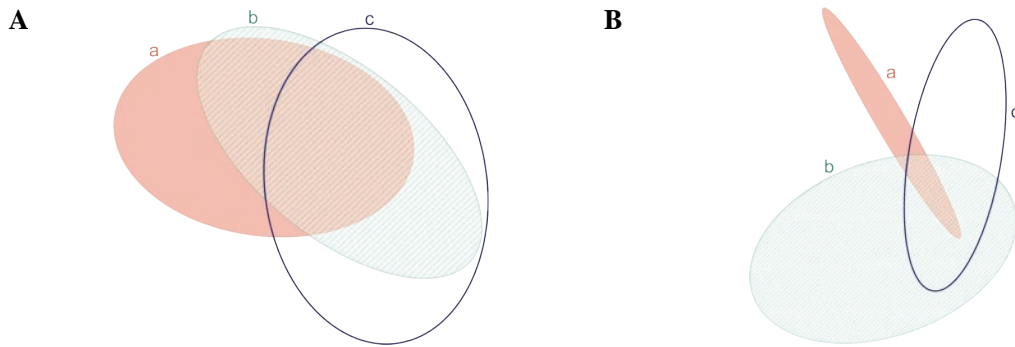
Figure 5.10: *Generated wellformed 3-Venn diagrams that are still included in a* DIAGLIB *type library.* Though these diagrams might not be considered aesthetically pleasing due to the shape of some of the zones, they are still included in the library (DIAGLIB_*design*) as they are wellformed 3-Venn diagrams. Zones *b* and *bc* in both A and B have irregular shapes. Zones *a* and *abc* have a very elongated shape.

The two generated libraries of this type are DIAGLIB_*design* and DIAGLIB_*eval*. The area specifications from these libraries were obtained by computing the zone areas of their diagrams using our analytic method M2 in Section 5.4.3. The area specifications obtained from DIAGLIB_*design* and DIAGLIB_*eval* were then used for our experiments at respectively the design phase (Section 5.6) and the evaluation phase (Section 5.7) of euler*APE* to assess when and how often local minima are encountered.

AREASPECLIB

A library of type AREASPECLIB provides a sample of the overall population of area specifications corresponding to the abstract description of a 3-Venn diagram, without any knowledge as to whether an accurate wellformed area-proportional 3-Venn diagram can be drawn with ellipses with respect to these area specifications.

Hence, a library of this type is made up of 10,000 random area specifications, each corresponding to the abstract description of a 3-Venn diagram with curves labelled *a*, *b* and *c*. Seven random real positive numbers are taken from a uniform distribution in the interval [1, 10000] to ensure a large diversity of zone areas. These numbers are then rounded up to two decimal places and each is randomly assigned to one of the seven zones to be located in at least one of the curves of the Venn diagram once it is drawn.

The two generated libraries of this type are AREASPECLIB_*design* and AREASPECLIB_*eval*. These libraries were used for our experiments at respectively the design phase (Section 5.6) and the evaluation phase (Section 5.7) of euler*APE*, to evaluate the number of area specifications from the overall population that can be drawn with ellipses and euler*APE* and compare these results with those of other methods, and to identify the types of area specifications for which an accurate area-proportional 3-Venn diagram can be drawn with ellipses.

5.5.2 The Area Specifications of the Four Generated Libraries

The area specifications obtained from DIAGLIB_*design* and DIAGLIB_*eval* are a special subset of the general population of area specifications, as an accurate area-proportional wellformed 3-Venn diagram with respect to these area specifications is known to exist. On the other hand, the area specifications in AREASPECLIB_*design* and AREASPECLIB_*eval* are representative of the general population, but it might not always be possible to draw an accurate diagram for some of these area specifications using ellipses.

The differences between the area specifications of these two library types is noted in Figure 5.11. These boxplots show the area of the zones in only one curve, the area of the zones in only two curves, the area of the zones in (exactly) the three curves, the total area of all the zones in only one curve, the total area of all the zones in only two curves, and the range of the zone areas, as percentages of the total area of all the zones in their respective area specification, for the 10,000 area specifications obtained from each of libraries DIAGLIB_*design*, AREASPECLIB_*design,* DIAGLIB_*eval* and AREASPECLIB_*eval*.
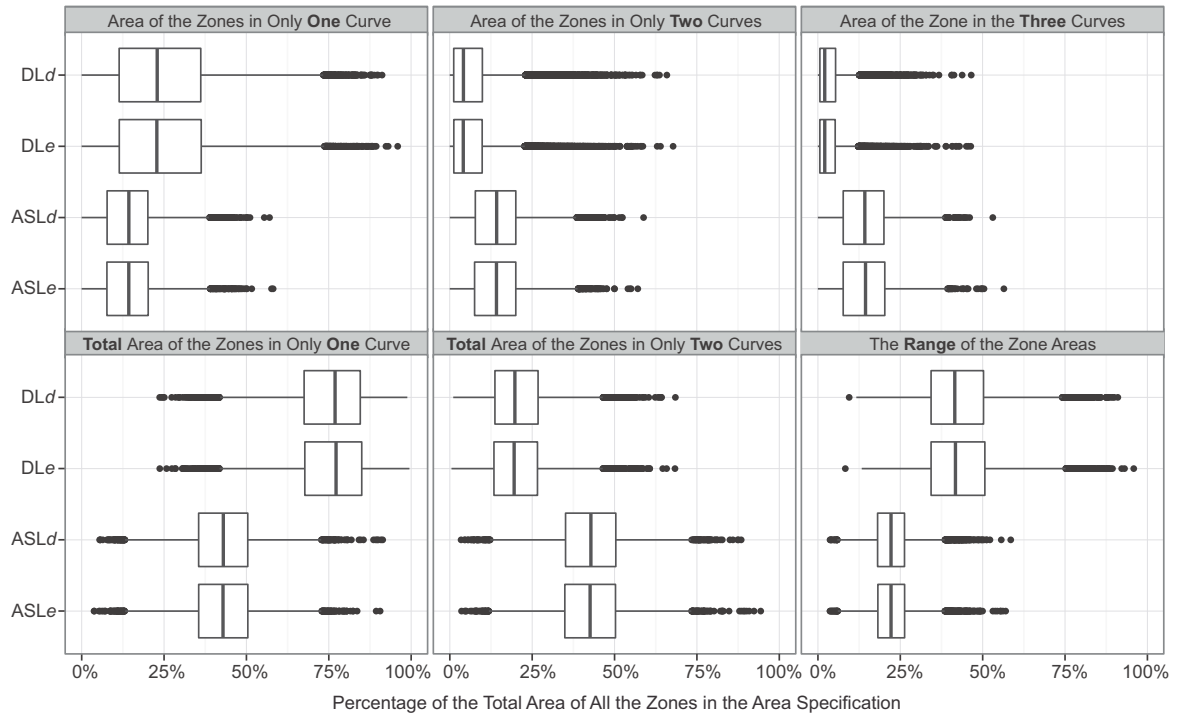
Figure 5.11: *The characteristics of the area specifications of the four generated libraries.* Libraries DIAGLIB_*design*, DIAGLIB_*eval*, AREASPECLIB_*design* and AREASPECLIB_*eval* are referred to as DL*d*, DL*e*, ASL*d* and ASL*e* respectively.

Looking at the plots for the area of the zones in only one and only two curves (the two leftmost top boxes in Figure 5.11) and those for the total areas of the zones in only one and only two curves (the two leftmost bottom boxes in Figure 5.11), we note that area specifications for which an accurate diagram is known to exist (obtained from a library of type DIAGLIB) are likely to have larger areas for the zones in only one curve (median 23.1%, mean 25.2%) than for the zones in only two curves (median 4.2%, mean 7.0%). The median zone area as a percentage of the total area of all the zones in the area specification was smaller (8.3%) for the area specifications obtained from the libraries of type DIAGLIB than that (14.3%) for the area specifications in the libraries of type AREASPECLIB. A median of 14.3% is equal to the mean zone area percentage (100% ÷ 7 zones = 14.3%). This indicates that the zone areas of the area specifications in AREASPECLIB type libraries are more evenly distributed than those obtained from DIAGLIB type libraries and are thus more representative of the general population for which an accurate diagram drawn with ellipses might not exist. As shown in Figure 5.11 (rightmost bottom box), the ranges of the zone areas were greater for the area specifications obtained from the DIAGLIB type libraries (median 41.6%, mean 43.3%) than those for the area specifications in AREASPECLIB type libraries (median 22.2%, mean 22.5%). These differences between the two library types suggest that not all area specifications can be depicted accurately with a wellformed Venn diagram drawn with ellipses. Characteristics of such area specifications are in Section 5.7.2. We now discuss our eulerAPE algorithm.

5.6 The Drawing Method – euler*APE*

Our drawing method euler*APE* is based on the simple hill-climbing optimization technique to generate an accurate area-proportional diagram using ellipses with respect to a given area specification. We explain the rationale for a heuristic and for choosing a simple hill-climber not other heuristics in Section 5.6.1.

The quantities provided in an area specification are first scaled as explained in Section 5.6.2 to ensure the same diagram is generated for different area specifications whose quantities are proportional. Later, the search for a solution that satisfies our goodness measure (Section 5.6.3) commences, so that a diagram with zone areas that are directly proportional to the scaled quantitative data of the area specification is generated. A cost function directs the optimization process to a good solution (Section 5.6.4), in that, starting with a rational diagram for the required zone areas (Section 5.6.5), the properties of the ellipses are adjusted based on the cost of the modified diagram (Section 5.6.6).

Design options were considered and evaluated, so pragmatic and experimental evidence supports the design of all the components of euler*APE* (details in the section corresponding to each component). A number of techniques were also adopted to avoid and handle local minima (Section 5.6.7).

5.6.1 The Rationale for a Simple Hill-Climber

Why a Heuristic

No analytic method can determine the properties of the curves of an area-proportional Venn diagram given only an area specification, since the zone areas are defined by the various properties of the curves enclosing the regions of the zones (Section 3.7.6 subsection 'The Need for a Heuristic'). Example, the area of the overlap between two ellipses is defined in terms of the centre, the semi-axes and the angle of rotation of the two ellipses. So irrespective of the representation and shape of the curves (circle, ellipse, polygon or any other shape), the function defining the zone areas is non-invertible [Chow, 2007].

Also, different diagrams with possibly different aesthetic features could be generated with respect to an abstract description (Section 3.4.3, Section 3.5; e.g., Figure 3.12 shows four ways how a 3-Venn diagram can be drawn—only Figure 3.12A is wellformed) and with respect to an area specification (e.g., the diagrams in Figure 5.2 and Figure 5.4 whose zone areas depict the same quantitative data).

Thus, a numerical method or a heuristic technique (as in Section 3.7.4) is required to adjust the curves' properties, in search for an adequate diagram whose zone areas are proportional to the required quantitative data and that satisfies any required aesthetic features, such as wellformedness as in euler*APE*.

Why a Simple Hill-Climber and Not Other Heuristics

Previous drawing techniques use different heuristic techniques (Section 3.7.4). Example the circle-based method `venneuler` [Wilkinson, 2012] uses the steepest descent method with an approximate gradient (i.e., local search optimization), while the polygon-based method VennMaster [Kestler et al., 2005, 2008] uses particle swarm and evolutionary optimization algorithms (i.e., global search optimization).

As noted by Wilkinson [2012] in his comparative analysis of `venneuler` to VennMaster, **stochastic optimization methods**, like those in VennMaster, use random variables and a random starting diagram, making the method non-deterministic as a different diagram is generated for the same area specification

each time VennMaster is run. Also, the randomly generated starting diagram might not be as adequate for the required quantitative data, so the optimization process might take long to converge to a solution, and it is more likely to reach a local minimum as indicated by studies (e.g., [Clark, 1976; Spence and Young, 1978]) for starting diagrams that are invariant or random. Wilkinson also argues that when a rational starting diagram is used, such that the required solution is close to the starting diagram, computationally expensive **global search optimization algorithms**, like those used by VennMaster and others such as **simulated annealing and genetic algorithms**, are not required. In fact, Wilkinson claims that though VennMaster took over 10 minutes to generate a diagram for some gene data, the *stress* value of the diagram (i.e., 0.036) was greater and thus worse than that of `venneuler`'s diagram (i.e., 0.014), which was generated in 10 seconds (when run on a MacBook Pro @2.5Ghz with 2GB RAM and Java Platform 1.5). Wilkinson provides various other cases where VennMaster fails to produce an adequate diagram, despite the use of global optimization and irrespective to whether the **particle swarm or the evolutionary algorithm** were used in the optimization. Also, generating a diagram in 10 seconds or less is preferred as it ensures the users' attention is maintained [Card et al., 1991; Miller, 1968]. However, this is not possible with **stochastic global search optimization methods** as they are notorious for a long search time (e.g., [Fouskakis and Draper, 2002; Mitchell et al., 1993; Talbi and Muntean, 1993]). VennMaster keeps the shape and orientation of the polygons unchanged and only optimizes their position, but its convergence time is still far from an immediate response.

Thus, a **local search optimization method** with a rational starting diagram that is close to the required solution should adequately draw area-proportional diagrams. The rational starting diagram ensures the algorithm is deterministic and reduces the likelihood of converging to the wrong solution or a local minimum. Yet, there are various local search optimization methods that can be used. Example, `venneuler` uses the steepest descent method, while the circle-based method by Chow and Rodgers [2005] uses a simple hill-climbing algorithm. In a **steepest descent method**, the gradient has to be computed and the minimization function has to be differentiated. The gradient is defined as a function of the accurate zone areas and so, computing this gradient is computationally expensive. In the case of `venneuler`, an approximation of the gradient is used, as the zone areas are not computed accurately with its quadrature numerical method. However, to compute a good approximation, various iterations involving repeated evaluation of its minimization function are required and thus, computing an approximation of the gradient could also be computationally expensive. If the approximate gradient is not close enough to the true gradient, the optimization diverges from the required solution.

The **hill-climbing algorithm** is simple and easy to implement. No gradient has to be computed and it has been preferred over other methods for various application areas (e.g., [Gent and Walsh, 1993; Guindon and Gascuel, 2003; Katoh and Toh, 2010; Storer et al., 1992]). The steepest descent method could aid handling ridges and speed up convergence [Hopgood, 2012]. However, it still unclear how problematic ridges are when drawing such diagrams. So, being the first drawing method that uses ellipses, **we decided to adopt a simple hill-climbing algorithm** which commences with a rational starting diagram, that is close to the final required solution and that has characteristics that could prevent convergence to a local minimum (Section 5.6.5), and adjusts the various properties of the ellipses systematically, accepting only the changes that minimize the cost function (Section 5.6.6). Though a simple technique, experimental evidence indicates that e*ulerAPE* rarely encounters a local minimum (Section 5.7.1) and when it does, the techniques we have adopted to handle local minima (Section 5.6.7) are able to find the required solution when one is known to exist. The time required to generate a diagram with e*ulerAPE* is similar to that of `venneuler` (Section 5.7.2) and so, considering that `venneuler`

uses a numerical method to compute the zone areas that is slightly faster than our analytic method (Section 5.4.4), the steepest descent method does not necessarily aid in speeding up convergence. It could also be that ridges are not much of a problem when drawing these diagrams and so, a simple hill climbing algorithm could be sufficient for our purpose. Also, using such a local search optimization, it is more likely for a diagram to be generated within the 10 second response time limit that ensures users' attention is maintained [Card et al., 1991; Miller, 1968] (as demonstrated in Section 5.7).

5.6.2 Scaling the Quantitative Data of the Area Specification

Given an area specification, the quantity assigned to each zone is scaled, before a diagram is drawn. The scaled quantity of a zone is then the required area of the zone in the diagram to be generated. In this way, the same diagram is generated for different area specifications whose quantities are proportional (e.g., the same diagram is generated for area specifications {$a=12$, $b=12$, $c=16$, $ab=4$, $ac=6$, $bc=12$, $abc=2$} and {$a=120$, $b=120$, $c=160$, $ab=40$, $ac=60$, $bc=120$, $abc=20$}).

So given an area specification, ω, of the form

$$\omega = \left\{ (a, k_1), (b, k_2), (c, k_3), (ab, k_4), (ac, k_5), (bc, k_6), (abc, k_7) \right\}$$

$$where \ \left\{ k_1, k_2, k_3, k_4, k_5, k_6, k_7 \right\} \subset \mathbb{R}_{>0}$$

the smallest quantity in ω, k_{min}, is found and each quantity k_i for $i \in \{1, \dots, 7\}$ is scaled to

$$k_i' = k_i \times \frac{C}{k_{min}} \quad where \ C \in \mathbb{N} \tag{5.15}$$

with each scaled value rounded up to 15 decimal places. So, the area specification with the scaled quantities, indicating the required zone areas in the diagram to be generated, is

$$\omega' = \left\{ (a, k_1'), (b, k_2'), (c, k_3'), (ab, k_4'), (ac, k_5'), (bc, k_6'), (abc, k_7') \right\}$$

$$where \ \left\{ k_1', k_2', k_3', k_4', k_5', k_6', k_7' \right\} \subset \mathbb{R}_{>0} \tag{5.16}$$

We set the constant C in Equation (5.15) to 100. Any other value for c could have been chosen. However, since the diagram display panel of eulerAPE is 720 pixels × 650 pixels, an area of 100 pixels^2 for the smallest zone in the diagram seems reasonable, so the smallest possible diagram (i.e., one whose zone areas are all equal) would have a total area of 700 pixels^2. The diagram displayed in eulerAPE's diagram display panel is a copy of the generated diagram with respect to ω', which is scaled (i.e., contracted or enlarged) to adequately fit in the panel.

Other scaling mechanisms have been considered and evaluated. For instance, the quantities assigned to the zones could be scaled by a factor that is inversely proportional to the quantity assigned to the zone in exactly the three curves. However, no difference in both quantity and time was noted and so, we decided to scale the quantities by a factor that is inversely proportional to the smallest quantity in the area specification as in Equation (5.15), to avoid handling very small zone areas.

5.6.3 The Diagram Goodness Measure

The main objective of eulerAPE is to generate wellformed area-proportional 3-Venn diagrams with regular, smooth and convex curves represented as ellipses and with zone areas that are accurate and directly proportional to the quantities in a given area specification (Section 5.2.3). We can easily check

whether the 3-Venn diagram satisfies all the wellformedness properties in Section 3.4.4 by computing the number of intersection points between every pair of ellipses. Each ellipse represents a set in the data provided in the area specification and so, the diagram always has unique curve labels and simple curves, satisfying wellformedness properties 2 and 3 in Section 3.4.4. For the other properties to be satisfied, so the diagram is wellformed, every pair of ellipses must intersect at only two points (as shown in example Figure 3.12A for a wellformed 3-Venn diagram). However, to check whether the zone areas are directly proportional to the quantities in a given area specification and thus accurate, a measure is required. We now discuss the measures we have considered and the measure we defined for eulerAPE.

Possible Measures

Various measures have been used by drawing methods (Section 3.7.4). `venneuler` [Wilkinson, 2012] uses its **statistical loss function *stress*** to measure the error in the diagram and how close it is to the required solution. The lower the *stress* value, the better the solution. Yet the *stress* value could be relatively low even if the diagram has missing zones (Section 3.7.4, Section 5.2.1). This is problematic as such diagrams do not depict all the required set relations and are thus more misleading than those with only inaccurate zone areas. So such a goodness measure is **not appropriate** for eulerAPE.

In **Chow and Rodgers**'s [2005] circle-based method, **an ad hoc measure** is used to determine whether the perceived order of the zones with respect to their size is consistent with that in the given area specification. If so, the diagram is good irrespective to whether the zone areas are proportional to the required quantities in the area specification, as the authors argue that readers are more concerned with the relative sizes of the zones (i.e., whether one zone is equal, larger or smaller than another) than the actual quantity they depict. The thresholds used to determine, example, that two zones are perceived equal in size are ad hoc, based on intuition rather than empirical evidence, as it is not known how zone areas are perceived and what differences in areas are not noted (Section 3.7.3). This measure does not meet our objective of generating diagrams with zone areas that are accurate and directly proportional to the quantitative data and thus, it is **not adequate** for eulerAPE.

The **goodness measure of a diagram in VennMaster** [Kestler et al., 2005, 2008] is primarily based on the absolute difference between the cardinality of the set intersection represented by a zone and the actual area of the zone scaled to correspond with the cardinality of the set intersection. It also includes a few weights to manage the diagram aesthetics. Example, errors of overlaps of large sets are weighted heavier than those of smaller sets. However, these weights are based on intuition rather than empirical evidence and conflict with our objective of generating zone areas that are all directly proportional to the required quantitative data. A measure that sums the absolute difference between the required and actual scaled zone areas is still **not appropriate**. Consider the following example. Let the area specification for which a diagram should be drawn be {a=1, b=1, c=1, ab=1, ac=1, bc=1, abc=1}. The sum of the absolute zone differences of a diagram d_1 whose zone areas correspond to {a=11, b=1, c=1, ab=1, ac=1, bc=1, abc=1} will be 10 and that of diagram d_2 whose zone areas correspond to {a=5, b=4, c=4, ab=1, ac=1, bc=1, abc=1} is also 10. However, d_1 could be less preferred than d_2 as one of the zones depicts a quantity that is 11 times as much as the required. If the maximum absolute zone differences is used, d_1 would have a value of 10, while d_2 a value of 4, indicating that d_1 is worst than d_2. Even so, it is difficult to determine an acceptable maximum absolute difference for a diagram to be good. A difference of 1 for a total area of the zones of 10000 is not the same as a difference of 1 for a total area of the zones of example 10.

A well-established measure known as the **cartographic error** [Dougenik, 1985], synonymous to the standard **relative error**, is used to determine the accuracy of the area of the regions of a cartogram in depicting the required quantitative data [Speckmann, 2006]. The error measure for each region is

$$\frac{|actualArea - requiredArea|}{requiredArea}$$

where *actualArea* is the area that the region has and *requiredArea* is the area that the region should have. The error of the cartogram is often the maximum or average error of all the regions [Speckmann, 2006]. This measure is still **not appropriate** for area-proportional Venn (or Euler) diagrams. We refer to this error measure as relative error whenever we are not referring to it in the context of cartography.

Firstly, the **relative error measure is not symmetric**. So, if the area specification for which a diagram should be drawn is ω_1 = {*a*=1, *b*=1, *c*=1, *ab*=1, *ac*=1, *bc*=1, *abc*=1} and the generated diagram d_1 has zone areas proportional to {*a*=10, *b*=1, *c*=1, *ab*=1, *ac*=1, *bc*=1, *abc*=1}, the relative error for zone *a* in d_1 is 9, so the maximum zone relative error in d_1 is 9 and the average 1.29. Yet if the area specification for which a diagram should be drawn is ω_2 = {*a*=10, *b*=1, *c*=1, *ab*=1, *ac*=1, *bc*=1, *abc*=1} and the generated diagram d_2 has zone areas proportional to {*a*=1, *b*=1, *c*=1, *ab*=1, *ac*=1, *bc*=1, *abc*=1}, the relative error for zone *a* in d_2 is 0.9, so the maximum zone relative error in d_2 is 0.9 and the average 0.13. Thus according to relative error, the error in d_1 with respect to ω_1 is much greater than the error in d_2 with respect to ω_2, even though in both d_1 and d_2 the only incorrect zone is *a* and the difference between the required and actual area of this zone *a* is 10 times as much. So an appropriate measure for area-proportional Euler diagrams must be symmetric, such that the error in d_1 and d_2 is treated equally.

Secondly, it is **difficult to determine a small value for relative error** for a diagram to be good. A cartogram with a maximum zone error in [0.01, 0.05] is often considered good and adequate in depicting the required quantitative data (e.g., [Buchin et al., 2012; van Kreveld and Speckmann, 2007]), as often a zero error is not feasible except for few special cases (e.g., [de Berg et al., 2010]). For area-proportional Euler diagrams, finding such a value is not so easy. Zones are in one or more curves. The area of all the zones and curves should be accurate and directly proportional to the required quantities. If a zone area is inaccurate, the area of the curve containing that zone would also be inaccurate. Quantities depicted by other zones can also be interpreted incorrectly, even if their area is proportional to the required quantity. This could occur as the zone areas are often compared with the area of the containing curve, of the other zones and of the entire diagram to understand the zone proportion and the depicted quantitative data.

Consider this example. The diagram in Figure 5.12A is accurate with respect to ω = {*a*=90, *b*=1, *ab*=1}. The relative error in each zone is zero. The diagram is clearly depicting the required cardinality for the sets and set intersections, as curve *a* is much larger than curve *b*, zone *ab* is half the size of curve *b*, and zones *ab* and *b* are similar in size. The diagram in Figure 5.12B is inaccurate with respect to ω as it corresponds to {*a*=90, *b*=10, *ab*=1}. Only zone *ab* has relative error greater than zero (i.e., 9), so according to this measure, only this zone is incorrect. Yet, looking at just Figure 5.12B, zones *ab* and *b* are clearly depicting different quantitative data, and despite the numeric labels, it is still unclear which is erroneous even when their areas are compared to that of zone *a*. In fact, all the curves and zones are erroneous. The area of curve *b* was meant to be 2.2% of that of curve *a*, but in Figure 5.12B, the area of curve *b* is 11.0% of *a*. The percentage area of zone *a*, *b*, and *ab* to the total diagram area was expected to be 97.8%, 1.1%, 1.1% respectively, but in Figure 5.12B, the percentage zone areas are 89.1%, 1.0%, 9.9% respectively. So all the zones are depicting an inaccurate percentage area of the total diagram, particularly zones *a* and *ab* with a similar percentage area difference of 8.7% and 8.8% respectively.

Yet relative error indicates that only zone *ab* is incorrect. This further shows the inappropriateness of such a measure to quantify the accuracy of area-proportional Euler diagrams with respect to an area specification. After all, cartograms and area-proportional Euler diagrams are different (Section 3.7.4) and a measure that is appropriate for cartograms is not necessarily appropriate for Euler diagrams.
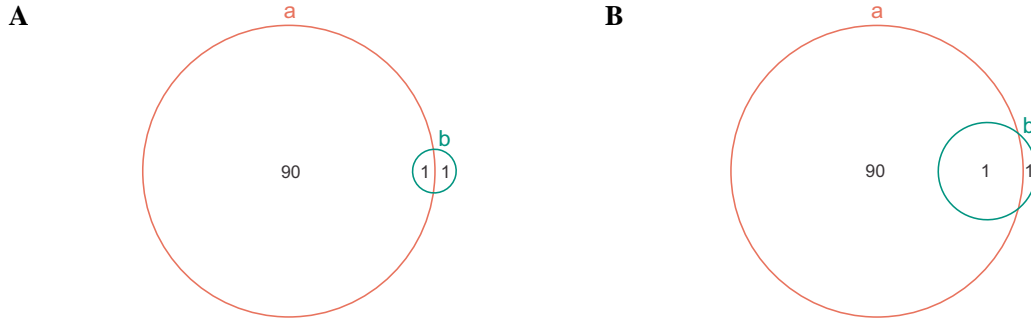
Figure 5.12: *A simple example to demonstrate that relative error is not adequate to measure the accuracy of an area-proportional Euler diagram with respect to an area specification.* A Venn diagram should be drawn for area specification ω = {*a*=90, *b*=1, *ab*=1}. The numeric labels indicate the quantity the zones should depict. (A) An accurate area-proportional Venn diagram with respect to ω. All zones have zero relative error. (B) An inaccurate area-proportional Venn diagram with respect to ω, corresponding to {*a*=90, *b*=10, *ab*=1}. Relative error zero for zones *a* and *b*, 9 for *ab*. So according to relative error, only zone *ab* is incorrect, despite that the percentage area of each zone to the total diagram area differs from that in A and ω.

Thus, we defined **our diagram error measure** based on the percentage area of each zone to the area of the total diagram. A similar measure was considered but not adopted by Chow and Rodgers [2005].

Our Measure

The accuracy of each zone area in depicting the required quantity should not be measured independently of the total diagram area. As shown in the previous section, an inaccuracy in one zone could make other accurate zone areas or curves in the diagram seem erroneous, as zones and curves might be compared for their area to be estimated. Instead of the absolute zone area, the zone area should be considered in respect to the total diagram area. This means that the proportion of each zone area to the total diagram area should be computed for both the required and actual diagram. The absolute difference of the two proportions for a zone would provide the error of that zone. The diagram error can then be equal to the largest error of the zones in the diagram. Thus, we define our diagram goodness measure as follows:

If

- ω is the area specification for which a diagram had to be drawn,
- *d* is a wellformed area-proportional diagram generated for ω,
- *Z* is the set of labels describing the required set of zones interior to the curves of the diagram,
- $\omega(z) \in \mathbb{R}^+$ is the quantity assigned to $z \in Z$ that should be depicted by the area of *z* in *d*, and
- $A(z) \in \mathbb{R}^+$ is the area of $z \in Z$ in *d*,

then the error in each $z \in Z$ is defined as

$$zoneError(z) = \left| \frac{\omega(z)}{\omega_S} - \frac{A(z)}{A_S} \right|$$

(5.17)

$$where \quad \omega_S = \sum_{z \in Z} \omega(z) \quad and \quad A_S = \sum_{z \in Z} A(z)$$

and the error in *d* is defined as

$$diagError = \max_{z \in Z}\left[zoneError(z)\right] \qquad (5.18)$$

so that *d* is a *good, accurate diagram with respect to ω*, in that its zone areas are accurately and directly proportional to the quantities in *ω*, if and only if

$$diagError \leq \varepsilon \quad where \ \ \varepsilon \rightarrow 0 \qquad (5.19)$$

In eulerAPE, $\varepsilon = 10^{-6}$, which value is consistent with that of other methods when defining a value for 0 in their implementation (e.g., venneuler [Wilkinson, 2012]). Identifying a value for *ε* is simpler for this measure than for others such as absolute or relative error, as the value of *zoneError* and *diagError* is always in the interval of [0,1]. In eulerAPE, the area of the zones in the diagram (i.e., $A(z) \in \mathbb{R}^+$ for every *z* ∈ *Z*) are computed using our analytic method M2 in Section 5.4.3.

We opted for a diagram error (*diagError*) that is defined as the maximum zone error rather than the average zone error, to ensure that the greatest error in the zones is considered and that the error in the other zones of the diagram is smaller or equal. The diagram error could also be defined as the sum of the error in all the zones. However, getting a diagram with the sum of all the zone errors less than or equal to *ε* where $\varepsilon = 10^{-6}$ is not feasible. If instead, $\varepsilon = 7 \times 10^{-6}$, then a diagram that is considered to be good when its zone errors are summed up, is also considered to be good by our measure in Equation (5.19) with $\varepsilon = 10^{-6}$. Also, in most areas, the maximum or the average error is considered (e.g., in cartography [Alam et al., 2012; Buchin et al., 2012; Speckmann et al., 2006; van Kreveld and Speckmann, 2007]). Thus, we decided to use the maximum zone error rather than the sum of the zone errors.

Before Equation (5.19) is used to determine whether a diagram accurately depicts the required quantities, the check explained at the beginning of this section (Section 5.6.3) is carried out to ensure that the diagram is wellformed. If the diagram is non-wellformed, then it is not a good solution as it does not satisfy our objectives (Section 5.2.3).

In contrast to relative error, our measure is symmetric. So, if the required area specification is $\omega_1 = \{a{=}1, b{=}1, c{=}1, ab{=}1, ac{=}1, bc{=}1, abc{=}1\}$ but the generated diagram d_1 corresponds to $\{a{=}10, b{=}1, c{=}1, ab{=}1, ac{=}1, bc{=}1, abc{=}1\}$, *diagError* in Equation (5.18) is 0.48. If, however, the required area specification is $\omega_2 = \{a{=}10, b{=}1, c{=}1, ab{=}1, ac{=}1, bc{=}1, abc{=}1\}$ but the generated diagram d_2 corresponds to $\{a{=}1, b{=}1, c{=}1, ab{=}1, ac{=}1, bc{=}1, abc{=}1\}$, *diagError* is also 0.48. So, the error in d_1 with respect to ω_1 and d_2 with respect to ω_2 is treated equally, as the error in both diagrams is in one zone whose quantity differs by a factor of 10 from the required. Thus, our measure is symmetric. This contrasts the asymmetric relative error measure which, as discussed earlier, returns a maximum error of 9 for d_1 with respect to ω_1 and 0.9 for d_2 with respect to ω_2, suggesting that the error in d_1 is greater than that in d_2, despite that both have one erroneous zone that differs by the same factor from the required.

Also, if we consider again the example we discussed earlier in relation to relative error and the diagrams in Figure 5.12 for area specification $\{a{=}90, b{=}1, ab{=}1\}$ where the zone areas of Figure 5.12A correspond to $\{a{=}90, b{=}1, ab{=}1\}$ and those of Figure 5.12A correspond to $\{a{=}90, b{=}1, ab{=}10\}$, we note that only Figure 5.12A satisfies Equation (5.19). This is so, as *zoneError*(*z*) in Equation (5.17) is 0 for every zone *z* in Figure 5.12A, but 0.087 (i.e., 8.7%), 0.001 (i.e., 0.1%) and 0.088 (i.e., 8.8%) for respectively zones *a*, *b* and *ab* in Figure 5.12B. So the diagram error (*diagError*) is 0 for Figure 5.12A and 0.088 for Figure 5.12B. Looking at only the quantities the zones of the diagrams relate to, zones *a* and *b* seem to be correct. However, due to the error in the area of zone *ab*, zone *a* and *b* seem erroneous

in Figure 5.12B, particularly when the zones are compared with one another, with a curve or with the area of the diagram, to gain better understanding of the proportions. Such effects can mislead the reader and so, these are noted by our zone error measure, so that *zoneError*(*a*) and *zoneError*(*b*) are greater than 0. Thus, in contrast to relative error, our zone error measure (*zoneError*) is capable of identifying zones that could appear erroneous and interpreted incorrectly due to errors in other zones.

Hence, if a generated diagram for an area specification is wellformed and satisfies Equation (5.19), the diagram accurately depicts the required quantitative data and is thus a good solution that meets our objectives (Section 5.2.3). Such a good solution would terminate the optimization algorithm discussed in Section 5.6.6. However, the cost function that directs the optimization process to this good solution is not based on Equation (5.19). We discuss this further in the next section (Section 5.6.4) and we provide pragmatic and experimental evidence to support this choice and the decision to devise a more specific cost function that aims at identifying and avoiding paths leading to local minima, while still directing the optimization to such a good solution. With a diagram goodness measure that it is independent of the cost function, it was possible to evaluate the effectiveness of different cost functions and still use the same measure to check the goodness of the diagram.

Though this work focuses on area-proportional 3-Venn diagrams drawn with ellipses, our diagram goodness measure Equation (5.19) can be used for any area-proportional Euler diagram with any number of curves that are not necessarily represented as ellipses.

5.6.4 The Cost Function

Our diagram goodness measure determines whether a wellformed area-proportional 3-Venn diagram is good, such that its zone areas are accurate and directly proportional to the quantities in a given area specification (Section 5.6.3). If a diagram is good, based on this measure, and its curves are regular, smooth and convex represented as ellipses, the diagram satisfies our objectives (Section 5.2.3). Such a good diagram would terminate the optimization, but to get to this diagram, during the optimization, other characteristics of the explored solutions have to be measured to determine and avoid any change that increases the likelihood of encountering a local minimum even though the error of the overall diagram is reduced. So during the optimization, a solution is preferred over another based on whether the cost to get to the good solution is reduced. Thus often the cost function that the optimization has to minimize differs in a few ways from the measure that is only used to determine if a diagram is good.

We now discuss the different functions that have been previously proposed and whether these could be considered for eulerAPE. We introduce the cost functions we devised, and we discuss our evaluation of the various functions we considered for eulerAPE, stating which of these was then employed and why.

Possible Functions

Various types of cost functions were used by previous drawing methods (Section 3.7.4). **venneuler** [Wilkinson, 2012] uses the same **loss function *stress*** as a diagram goodness measure and as a function to direct the optimization. Using this function, generated diagrams with missing zones could still have a relatively low *stress* value (Section 3.7.4, Section 5.2.1) despite that such diagrams are more misleading than ones with only inaccurate zone areas. However, during our optimization, no diagram that is not a 3-Venn diagram is allowed and so, ***stress* could still be considered** for eulerAPE.

Chow and Rodgers's [2005] **fitness function** was based on their ad hoc goodness measure, which as discussed earlier in Section 5.6.3 subsection 'Possible Measures' does not meet our objectives and should not be considered for eulerAPE. However, the authors also provide **an 'idealistic' function**, which they did not use nor evaluate. The function is the sum of two components. The first is the sum of the squared *zoneError* Equation (5.17) for all the required zones, while the second is the summation of an inequality that verifies whether the ordering of the zones with respect to their area is consistent with that defined in the area specification. If the first component is minimized to a value very close to zero, all the zone areas would be directly proportional to the required quantitative data and the second component would not be required. The intent for this second component was namely to help the algorithm generate less misleading diagrams with circles, since circles cannot draw accurate diagrams for most 3-set data.

Hence, since the **first component of the 'idealistic' function** by Chow and Rodgers [2005] is based on our *zoneError* Equation (5.17), it is worth considering this component (without the second) as a cost function for eulerAPE. Such a function would be more effective in directing our optimization than if our *diagError* Equation (5.18) was used. This is so, as squaring the *zoneError* could aid the optimization in identifying inappropriate changes to the diagram as large errors are weighted more heavily than smaller ones. Also, squaring would ensure that if, example, a diagram has only one zone with *zoneError* > 0 and its *zoneError* is 1, and another diagram has two zones with *zoneError* > 0 and both have *zoneError* 0.5, the cost of the first would be greater than that of the second. If the cost was the sum of *zoneError*, both diagrams would have the same cost, even though the first is less preferred than the second. This is in fact one of the ways how the cost function often differs from the goodness measure.

Also, it is typical for the cost to be the sum or mean rather than the maximum of the squared error, so the error of all the zones is considered. If the maximum of the squared error is considered and the optimization has to choose which of the diagrams between d_1 and d_2 reduces the cost, where d_1 has one zone with *zoneError* 1 and the rest have *zoneError* 0, and d_2 has two zones each with *zoneError* 1 and the rest have *zoneError* 0, the optimization would not be able to choose between d_1 and d_2 as both would have the same cost despite that d_2 is worse. For the diagram goodness measure Equation (5.19), considering the maximum error in the zones is sufficient as the measure determines whether the diagram error is less than or equal to a value close to zero. So the worst possible diagram considered good is one whose *zoneError* for all the zones is less than or equal to a value close to zero (in eulerAPE, 10^{-6}). When the cost is the sum of the squared errors, large errors would dominate the final cost, so the largest errors in the zones are minimized first. Such large errors in the zones still affect the mean and so, often, when Euler diagrams with any number of zones are drawn, the mean of the squared error is used, as diagrams with the same zone errors but with a different number of required zones would get a different cost if the sum is used. Since this work focuses on 3-Venn diagrams, the area specifications will always have the same number of zones, so the **sum or mean of the squared** *zoneError* could be considered for eulerAPE.

The **cost function used in the polygon-based method VennMaster** [Kestler et al., 2008] tries to direct the optimization to a solution that is a compromise between accuracy and aesthetics (discussed earlier in Section 5.6.3 subsection 'Possible Measures'). Thus the zone areas of the targeted solution are not necessarily directly proportional to the depicted quantity. This conflicts with our accuracy objective, so this cost function cannot be considered for eulerAPE. It is however interesting that the function heavily weights overlaps that are unwanted or missing, in a bid to generate only the required zones.

Cost functions based on cartographic error [Dougenik, 1985] are often used to draw accurate cartograms [Speckmann, 2006]. Example, when evolutionary strategies were used, the part of the cost

function that determines the error of the cartogram's region areas was defined as the mean of the squared cartographic error [Buchin et al., 2012]. Another method that adopted a linear programming approach used the sum of the squared cartographic error [Speckmann et al., 2006]. As a goodness measure, the maximum or average cartographic error of the cartogram's regions is used, but as a cost function, the mean or sum of the squared cartographic error is used, for the same reasons discussed earlier for our *zoneError*. This cartographic error or standard relative error is not symmetric, so as shown earlier in Section 5.6.3, it cannot be used as a goodness measure of area-proportional Euler diagrams and more so, it cannot be used to define a cost function for these diagrams. Example, consider that two diagrams d_1 and d_2 have only one zone area not equal to the required, and in d_1 the zone is 10 times larger than required and in d_2 the zone is 10 times smaller than required. The relative error of the zone in d_1 is 9 and that of the zone in d_2 is 0.9. So the cost of d_2 would be less than that of d_1 (e.g., if the cost function is the sum of the squared relative error than the cost of d_1 is 81 and the cost of d_2 is 0.81) and thus, during the optimization, d_2 would be preferred over d_1 despite that both diagrams are equally wrong. Yet relative error is well-established and it is worth evaluating **a cost function based on relative error** for eulerAPE.

In an informal experimentation, we observed that when most of the cost functions discussed above are used, including venneuler's *stress* and Chow and Rodgers's 'idealistic' function, a local minimum is often reached when the optimization is driven to a path that reduces the overall diagram error at the expense of diminishing a zone that is meant to be larger, up to a point when the zone is close to non-existent and the optimization stops. Thus, **we devised novel cost functions** that drive the optimization to a good diagram Equation (5.19) and yet, prevent it from taking such paths. We now discuss these cost functions and our experimental comparison of these and other functions discussed here for eulerAPE.

Our Novel Functions

As demonstrated in the previous section, the cost function often differs from the goodness measure in a number of ways, as the former determines whether a change in the diagram will direct the optimization to a path that leads to a good solution, while the latter only determines whether a diagram is good. So, besides checking if a change reduces the overall error in the diagram, the cost function should also determine and prevent the optimization from taking paths that lead to a local minimum.

We observed that a local minimum is often reached when the error of the diagram is reduced at the expense of one zone diminishing, even though it should be larger, up to a point when it is barely visible. To prevent the optimization from taking such paths, we devised the following cost function, so that:

If

- ω is the area specification for which a diagram should be drawn,
- ω' is the area specification with the scaled quantities of ω obtained as defined in Section 5.6.2 indicating the required zone areas in the required good diagram for ω,
- d is a wellformed area-proportional diagram that is explored for ω during the optimization,
- Z is the set of labels describing the required set of zones interior to the curves of the diagram,
- $\omega'(z) \in \mathbb{R}^+$ is the area that $z \in Z$ should have in the required good diagram, and
- $A(z) \in \mathbb{R}^+$ is the area of $z \in Z$ in d,

then the cost of d is defined as

$$\frac{1}{|Z|} \sum_{z \in Z} \frac{(\omega'(z) - A(z))^2}{A(z)} \tag{5.20}$$

In eulerAPE, the zone areas ($A(z) \in \mathbb{R}^+$ for each $z \in Z$) are computed by our algorithm M2 in Section 5.4.3.

Hence, **the cost of a diagram is the mean of the cost of all the zones in that diagram**. The use of the sum or mean of the cost of all the zones is typical and preferred over maximum, as explained in the previous section. This work focuses on 3-Venn diagrams and so, the sum could be used instead of the mean as the number of zones and thus $|Z|$ will always be seven. However, we opted for the mean so Equation (5.20) could be used as a cost function for other future algorithms that are devised to draw Euler diagrams with any zones. Yet its effectiveness to handle such diagrams still needs to be evaluated.

To compute the error of the zone areas, we consider **the squared absolute difference between the required and actual zone area (i.e., $(\omega'(z) - A(z))^2$ for $z \in Z$)**. We square the difference for the reasons discussed in the previous section, but primarily, to heavily weight large errors, as shown in Figure 5.13 where, as an example, the value of $(\omega'(z) - A(z))^2$ for $z \in Z$ is shown for $A(z)$ in [0.01, 10] and for $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$. Squaring the difference is also typical when defining cost functions. Since we scale the required area for each zone in the given area specification, such that the quantities in ω' are the zone areas in the required diagram, it is adequate to consider the absolute zone areas. This contrasts our diagram goodness measure, which considers the zone areas relative to the diagram area, as an error in one zone makes other accurate zones or curves seem erroneous (Section 5.6.3). Yet its objectives are different from those of the cost function, as it only determines whether a diagram is good and whether the optimization should be terminated. The cost function, on the other hand, determines whether a change in the diagram will aid the optimization reach a good solution and so, if it considers relative zone areas as in our goodness measure, it could be too restrictive, particularly at the initial stages of the optimization when major changes are explored in search for a good yet non-refined solution. The area of most of the zones in such initial solutions will not be equal to that required and so, considering the error of each zone independently of one another could be more effective, particularly in terms of time and the number of iterations to get to a good solution. This is possible by considering the absolute difference between the required and actual zone area. If this difference is minimized to a value close to zero for all the zones, the obtained solution is still likely to meet our diagram goodness measure. VennMaster's [Kestler et al., 2008] cost function, heavily weights overlaps that are unwanted or missing, but it does not attempt to identify and prevent the optimization from taking paths that lead to such diagrams.
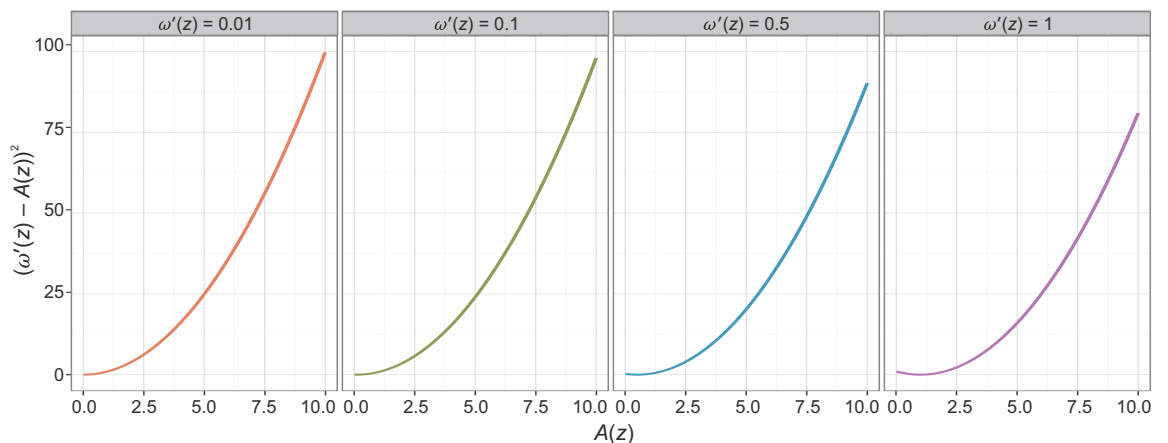
Figure 5.13: *The cost of a zone z when it is computed as $(\omega'(z) - A(z))^2$ where $\omega'(z)$ and $A(z)$ are respectively the required and actual area of z—this example illustrates the cost for A(z) in [0.01, 10] and for $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$*. This is how the cost of every zone in the diagram would be computed if only the squared absolute difference between the required and actual zone areas is considered.

The **denominator $A(z)$ for $z \in Z$** in our function prevents the optimization from taking paths that would reduce the overall diagram error, at the expense of constantly diminishing one zone area that should be larger, up to the point when the zone is close to non-existent and a local minimum is reached. With our function, a zone with a very small area compared to the required in ω' will have a very large cost to prevent the optimization from taking such paths. This is illustrated in Figure 5.14 where, as an example, the cost of a zone $z \in Z$ as computed by our cost function Equation (5.20) is shown for $A(z)$ in [0.01, 10] and $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$. The larger the absolute difference between the required and actual area, and the smaller the actual zone area, the greater the zone cost. Yet, if the absolute difference is less than one, which is often the case when the required good diagram is about to be reached, and the actual zone area is also less than one, then since the absolute difference is squared, the cost will not be too large and the optimization will proceed with this path. So, if the absolute difference between the required and actual zone area is 10^{-6} (a value which eulerAPE considers very close to zero) and the actual zone area is less than one (possibly close to zero like 10^{-6}), the zone cost would still be close to zero (i.e., 10^{-6} or smaller). If, however, the actual zone area is greater than one, the influence of the squared absolute difference between the actual and required zone area on the zone cost is reduced. This should not be problematic as by squaring the absolute differences, large differences are heavily weighted, as shown in Figure 5.13. Weakening these heavy weights on outliers could in fact be beneficial.

Consider, for instance, two zones z_1 and z_2 where $A(z_1) = 20$, $\omega'(z_1) = 10$, $A(z_2) = 2$, $\omega'(z_2) = 1$. The actual area of both z_1 and z_2 is twice as large as that required. If the cost of the zone was $(\omega'(z) - A(z))^2$ for $z \in Z$, the cost of z_1 would be 100 and that of z_2 1, meaning that though both z_1 and z_2 are twice as large as required, z_1 would have a much greater impact on the final diagram cost than z_2. However, if the zone cost is $(\omega'(z) - A(z))^2 / A(z)$ as in Equation (5.20), the cost of z_1 would be 5 and that of z_2 0.5. So by considering the squared absolute difference relative to the actual zone area, the cost of all the zones would get a fair chance to contribute to the total diagram cost independently of the size of the required zone area. As illustrated in Figure 5.13 and Figure 5.14, as the actual zone area increases, the cost of the zone increases much rapidly when the squared absolute difference is considered (Figure 5.13) than when the squared absolute difference relative to the actual zone area is taken into account (Figure 5.14).
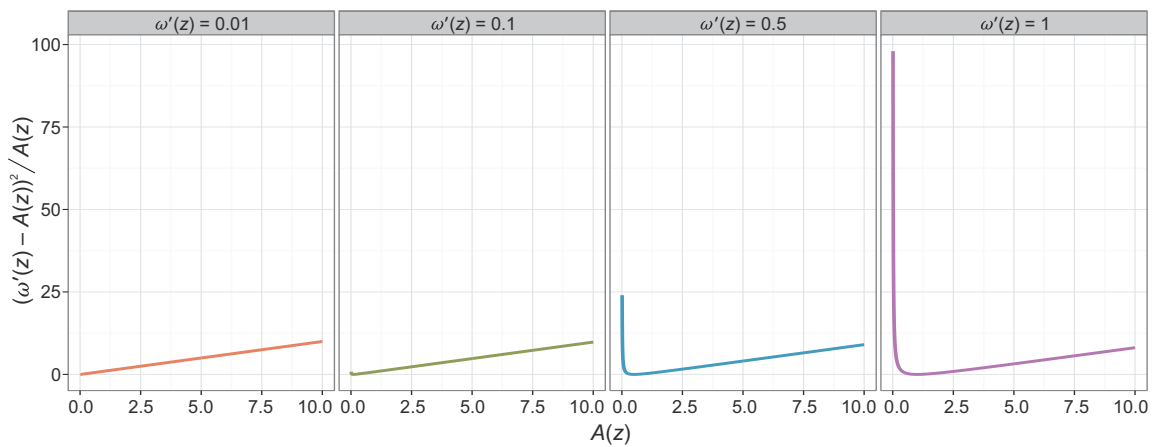
Figure 5.14: *The cost of a zone z when it is computed as $(\omega'(z) - A(z))^2 / A(z)$ where $\omega'(z)$ and $A(z)$ are respectively the required and actual area of z—this example illustrates the cost for $A(z)$ in [0.01, 10] and for $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$.* This is how our non-dimensionless cost function in Equation (5.20) computes the cost of every zone in the diagram.

Our **cost function in Equation (5.20) is non-dimensionless**, but as explained in Section 5.6.2, the quantities in the given area specification ω are always scaled using a predefined constant (hardcoded in eulerAPE) and made available as ω' before the optimization commences. So the area of a zone $z \in Z$, $A(z)$, in a good diagram generated for ω should be equal to $\omega'(z)$. Such scaling was also used in VennMaster [Kestler et al., 2008]. The diagram displayed in eulerAPE's panel is a copy of the diagram generated for the required zone areas in ω', which is enlarged or compressed to adequately fit in the panel.

By scaling the given area specification, our non-dimensionless cost function can be used. Yet we still considered **ways how to make Equation (5.20) dimensionless**. The denominator $A(z)$ for $z \in Z$ could be squared as in

$$\frac{1}{|Z|} \sum_{z \in Z} \left(\frac{\omega'(z) - A(z)}{A(z)} \right)^2 \tag{5.21}$$

or the numerator could be the absolute difference between the required and actual zone area not squared as in

$$\frac{1}{|Z|} \sum_{z \in Z} \frac{|\omega'(z) - A(z)|}{A(z)} \tag{5.22}$$

By squaring the denominator, the cost of a zone z in Equation (5.21) would be much smaller than that in Equation (5.20) when $A(z) > 1$ and much greater when $A(z) < 1$. This would have a different effect from that intended by our non-dimensionless function in Equation (5.20), as large errors would be less easily identified and the actual zone area would have a greater impact on the cost from that intended to aid identify zones that are diminishing in size when they should not. As an example, Figure 5.15 shows the cost of a zone $z \in Z$ computed by Equation (5.21) for $A(z)$ in [0.01, 10] and $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$.
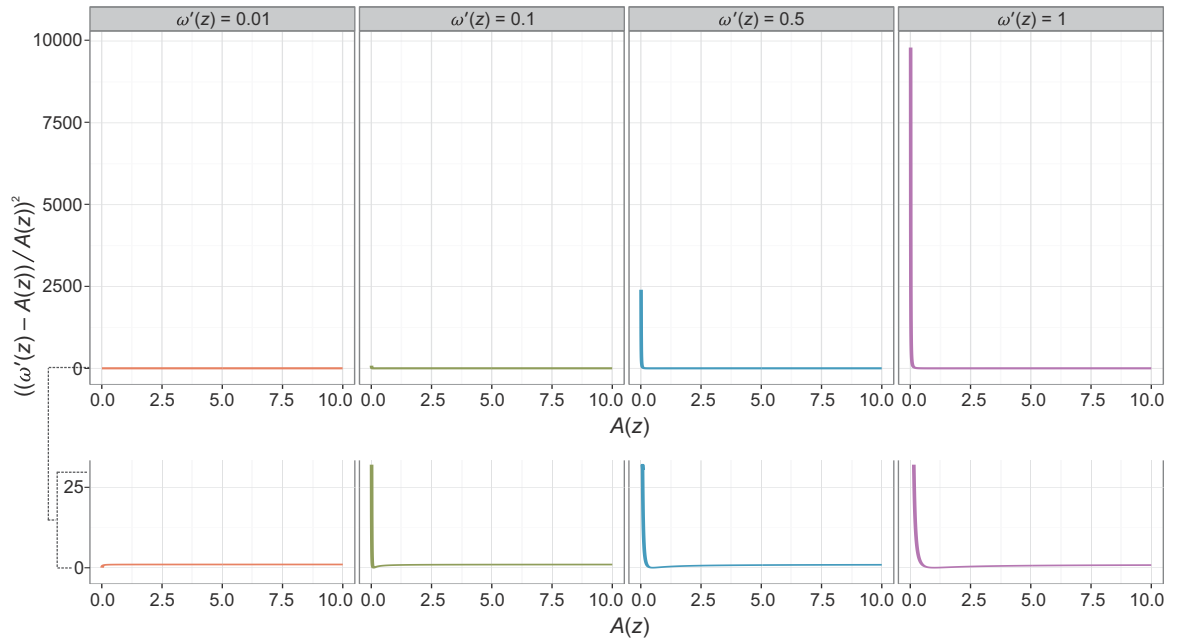
Figure 5.15: *The cost of a zone z when it is computed as $((\omega'(z) - A(z))/A(z))^2$ where $\omega'(z)$ and $A(z)$ are respectively the required and actual area of z—this example shows the cost for A(z) in [0.01, 10] and for $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$. This is how our dimensionless cost function in Equation (5.21) computes the cost of every zone in the diagram. Below is a magnified view of the plots for the cost in [0, 25].*

The zone cost in Equation (5.22) is the same as that in Equation (5.21) except that it is not squared. So if large zone errors result in a smaller cost in Equation (5.21) than in Equation (5.20), in Equation (5.22)

the cost would be even smaller, making it harder to identify inappropriate paths. Also, if the absolute difference between the actual and required zone area is less than one and the actual zone area is greater than or equal to one, then the zone will have a larger cost in Equation (5.22) than in the other two functions, making it even harder to determine which path leads to a good solution. Thus, the reason why typically a cost function squares the error, even when the related goodness measure only considers the absolute error (e.g., cartographic error). As an example, Figure 5.16 illustrates the cost of a zone $z \in Z$ as computed by Equation (5.22) for $A(z)$ in [0.01, 10] and for $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$. So we expect **Equation (5.21) and Equation (5.22) to be less effective than Equation (5.20) and for Equation (5.22) to perform worse than Equation (5.21)**. Yet it is worth evaluating these dimensionless cost functions.
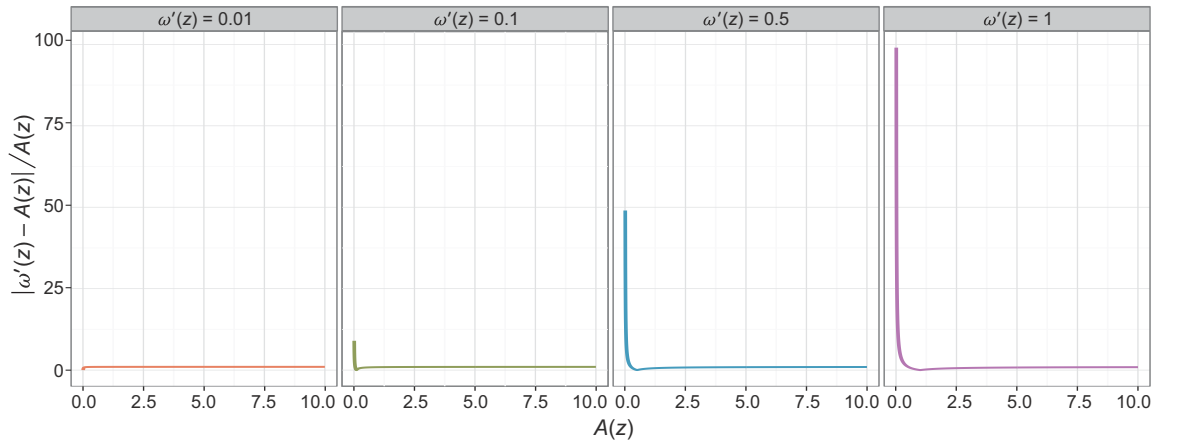
Figure 5.16: *The cost of a zone z when it is computed as $|\omega'(z) - A(z)|/A(z)$ where $\omega'(z)$ and $A(z)$ are respectively the required and actual area of z—this example illustrates the cost for A(z) in [0.01, 10] and for $\omega'(z) \in \{0.01, 0.1, 0.5, 1\}$. This is how our dimensionless cost function in Equation (5.22) computes the cost of every zone in the diagram.*

We will now discuss our evaluation and experimental comparison of these cost functions and others discussed in the previous section that are worth considering for eulerAPE.

Evaluation of Possible Functions

To choose the most effective cost function for eulerAPE, we conducted an experimental comparison of the following cost functions discussed in the previous sections. So, if

- ω is the area specification for which a diagram should be drawn,
- ω' is the area specification with the scaled quantities of ω obtained as defined in Section 5.6.2 indicating the required zone areas in the required good diagram for ω,
- d is a wellformed area-proportional diagram that is explored for ω during the optimization,
- Z is the set of labels describing the required set of zones interior to the curves of the diagram,
- $\omega'(z) \in \mathbb{R}^+$ is the area that $z \in Z$ should have in the required good diagram, and
- $A(z) \in \mathbb{R}^+$ is the area of $z \in Z$ in d computed using our analytic method M2 in Section 5.4.3,

then the evaluated cost functions were:

- **F1**. The statistical *stress* function used in `venneuler` [Wilkinson, 2012] that was computed with the source code of `venneuler` version 1.1-0 (provided to us by the author) but using $\omega'(z)$ and $A(z)$ for the required and the actual area of a zone $z \in Z$;

- **F2**. The first component of the 'idealistic' function of Chow and Rodgers [2005], based on *zoneError* Equation (5.17) Section 5.6.3, which we computed as

$$\sum_{z \in Z} (zoneError(z))^2 \;=\; \sum_{z \in Z} \left(\frac{\omega(z)}{\omega_S} - \frac{A(z)}{A_S} \right)^2 \; where \; \omega_S = \sum_{z \in Z} \omega(z) \; and \; A_S = \sum_{z \in Z} A(z) \;;$$

- **F3**. The sum of the squared relative error of the zones, which we computed as

$$\sum_{z \in Z} \left(\frac{\omega'(z) - A(z)}{\omega'(z)} \right)^2 \;;$$

- **F4**. The maximum of the relative error of the zones, which we computed as

$$\max_{z \in Z} \left[\frac{|\omega'(z) - A(z)|}{\omega'(z)} \right] \;;$$

- **F5**. The sum of the relative error of the zones, which we computed as

$$\sum_{z \in Z} \frac{|\omega'(z) - A(z)|}{\omega'(z)} \;;$$

- **F6**. Our non-dimensionless cost function Equation (5.20) defined in the previous section as

$$\frac{1}{|Z|} \sum_{z \in Z} \frac{(\omega'(z) - A(z))^2}{A(z)} \;;$$

- **F7**. Our dimensionless cost function Equation (5.21) defined in the previous section as

$$\frac{1}{|Z|} \sum_{z \in Z} \left(\frac{\omega'(z) - A(z)}{A(z)} \right)^2 \;;$$

- **F8**. Our dimensionless cost function Equation (5.22) defined in the previous section as

$$\frac{1}{|Z|} \sum_{z \in Z} \frac{|\omega'(z) - A(z)|}{A(z)} \;.$$

As discussed in the previous sections, since this work focuses on 3-Venn diagrams, the diagrams and the area specifications will always have the same number of zones and so, independently of whether the sum or the mean is used to collate the cost of every zone to the cost of the diagram, the optimization will be directed to the same solution. Thus, we evaluated F2 and F3, but not variants of them using the mean. Our cost functions F6, F7 and F8 use the mean as they were devised to handle Euler diagrams with any number of curves. For this work and this experimentation, we could have used variants that use the sum, but we did not for integrity, to evaluate the functions as they were devised. Though both F3 and F4 are based on relative error, in areas such as cartography F4 is used as a goodness measure, while F3 is used a cost function. We expect F4 to perform worse than F3 for various reasons discussed earlier, primarily for not squaring the cost of the zones and for only considering the maximum cost of the zones. However, relative error has never been adopted for the generation of area-proportional Venn or Euler diagrams and so, we decided to evaluate F3 and F4, as well as F5 that is based on the sum rather than the maximum cost of the zones.

Each cost function was used by the optimization algorithm discussed in Section 5.6.6 to generate diagrams for the two sets of 10,000 area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design* (Section 5.5). Throughout this section, these two sets of area specifications are respectively referred to **DL*as*** and **ASL*as***. To conduct the experiment, the cost function Equation (5.20) in step 8 of the optimization algorithm (Algorithm 5.3) was replaced by any one of the above that was being evaluated. Independently of the cost function, the optimization terminated when a good diagram satisfying Equation (5.19) was found or when a local minimum was reached, as in Algorithm 5.3. In this way, the cost function only determined whether a change in the diagram leads the optimization to a good solution. This also ensured integrity in our evaluation, such that the definition of a good diagram for a given area specification was consistent between cost functions. The rerun option in our optimization algorithm was disabled, so the optimization was always run once with one set of values for its parameters. The error of the generated diagrams was measured using our *diagError* Equation (5.18) and the error was then used to determine how many of the generated diagrams satisfied our diagram goodness measure Equation (5.19). The number of iterations and the time taken to generate the diagrams was also recorded. All the experiments were run on an Intel Core i7-3770 CPU @3.4GHz with 8GB RAM, 64-bit Microsoft Windows 7 Professional SP1 and Java Platform 1.7.0_10.

In this evaluation, **we aimed to**

- identify which cost function is most effective in avoiding local minima and in driving the optimization to a good diagram when one is known to exist for the given area specification, and

- identify which cost function is most effective in driving the optimization to a good diagram when it is not known whether such a diagram exists for the given area specification.

For the first objective, we compared the number of generated diagrams for DL*as*, that satisfied our diagram goodness measure Equation (5.19). Libraries of type DIAGLIB like DIAGLIB_*design* are a collection of wellformed 3-Venn diagrams with random ellipses (Section 5.5.1). The zone areas of these diagrams are computed, so a collection of area specifications for which a good diagram is known to exist is defined. Thus, failing to generate a good diagram for such area specifications indicates that a local minimum was reached. For the second objective, we compared the number of generated diagrams for ASL*as* that satisfied our diagram goodness measure Equation (5.19), as AREASPECLIB type libraries like AREASPECLIB_*design* consist of area specifications with random quantities for which a wellformed accurate area-proportional 3-Venn diagram drawn with ellipses might not exist (Section 5.5.1).

We hypothesized that our non-dimensionless cost function F6 would be the most effective in generally directing the optimization to a good diagram, as it was devised to identify and avoid paths that lead to a local minimum, particularly those that reduce the overall diagram error at the expense of one zone diminishing that should be much larger, up to a point when it is barely visible and the optimization has to terminate. We expected our dimensionless variants F7 and F8 to be less effective than F6 for the reasons discussed in subsection 'Our Novel Functions' and also, possibly less effective than F1 and F2 in reaching a good solution, as the required adaptations to make our cost function dimensionless make inappropriate paths less easily identified. Even so, F7 and F8 will not lead to diagrams whose overall error is reduced at the expense of diminishing one zone that should be larger. We hypothesized that F2 will require more iterations and time to get to a good diagram than F6, as F2 considers the zone areas relative to the area of the diagram rather than the absolute zone areas independently of the other zones as in F6 (subsection 'Our Novel Functions'). Relative error is not entirely appropriate for drawing these

diagrams (subsection 'Possible Functions') and so, we expected F3 to perform worse than F1, F2 and F6, and for F4 and F5 together with F8 to be the least effective as the error in the zones is not squared.

The **experimental results** are available in Appendix A.

This **evaluation and experimental comparison indicates** that our non-dimensionless cost function F6 is the most effective in:

- Generating good diagrams for area specifications for which a good diagram is known to exist (Appendix A.2.1);

- Converging to diagrams that have a low *diagError* even when a good diagram cannot be drawn with respect to the given area specification (Appendix A.2.2);

- Identifying and avoiding paths that lead the optimization to a local minimum when the overall error of the diagram is reduced at the expense of diminishing a zone to a point where it is close to non-existent and its actual-to-required zone area ratio is close to zero (Appendix A.2.3);

- Taking the least amount of the time and iterations to generate the diagrams, particularly for area specifications for which a good diagram is known to exist (Appendix A.2.4);

- Generating a large majority of the diagrams (97.3%, $N = 20000$) within a time (1 second) that ensures that the users' train of thought is maintained, and generating nearly all the diagrams (99.6%, $N = 20000$) within a time (10 seconds) that ensures that the users' attention is maintained (Appendix A.2.4 "Response Time Effectiveness").

The effectiveness of F6 over the other cost functions with respect to the generation of good diagrams, time and the number of iterations was highly evident for DL*as*, but less evident for ASL*as*. No statistically significant differences were revealed between the percentages of good diagrams of F1, F2 and F6 for ASL*as*. A small significant effect between F6 and both of F1 and F2 on time and number of iterations for ASL*as* was reported. However, though the medians of F1 and F2 were marginally smaller than those of F6, the means of F6 were much lower than those of F1 and F2.

This evaluation also demonstrated that all of the cost functions (F1-F5), except for those we devised (F6-F8), often direct the optimization to a local minimum as the actual-to-required zone area ratio is reduced to a value close to zero. So F1-F5, all of which had been previously used or proposed should be avoided. These cost functions include: (F1) *stress* that is used by `venneuler` [Wilkinson, 2012], which is considered the current most effective circle-based method; (F2) Chow and Rodgers's [2005] 'idealistic' function, that is a variant of *diagError* and our goodness measure; (F3, F4, F5) the cost functions based on relative error, that are often effective in drawing accurate cartograms. This also indicates that a cost function differs from the diagram goodness measure, as it should reduce the diagram error, but also identify and avoid paths leading to a local minimum. The latter is possible by a cost function that heavily weights zones whose area is very small with respect to that required, as done in F6-F8.

Following F6, cost functions F2 and F1 were the most effective in generating a good diagram. Next there were F7 and F3 and finally, F4, F5 and F8, which functions led to the generation of very few good diagrams as they do not square the error in the zones, making inappropriate paths less distinguishable. As expected, our dimensionless cost functions F7 and F8 were not as effective as our non-dimensionless F6 and less effective than F1 and F2. However, both achieved results (in terms of *diagError*, time and iterations) similar to those of functions based on relative error, namely F3 which squares the cost of the zones as our F7, and F5 which sums but does not square the cost of the zones as our F8.

For 0.5% of the area specifications for which a good diagram was known to exist (i.e., DL*as*), F6 reached a local minimum and generate an inaccurate with $diagError \leq 1.85 \times 10^{-2}$. As we demonstrate in Section 5.7.1, these very few cases are handled by the rerun option in our algorithm, which reruns the optimization using different starting values for its parameters. However, this option was disabled in this experiment to determine the number of good diagrams generated by only the first run.

The Function Employed By euler*APE*

As demonstrated, our specially devised non-dimensionless cost function in Equation (5.20) (F6 in our evaluation) has various characteristics that make it more effective than other cost functions, in directing the optimization to a wellformed diagram whose zone areas are accurate and directly proportional to the quantities in a given area specification, and in identifying and avoiding paths that lead to a local minimum, particularly in those cases when the overall error of the diagram is reduced at the expense of diminishing the area of one zone. Thus, cost function Equation (5.20) was employed by euler*APE*.

We now discuss the methods we devised and evaluated to generate a rational starting diagram that is close to the required diagram and that has characteristics that could further aid avoid local minima.

5.6.5 The Starting Diagram

The optimization process has to start with a solution. This is often an arbitrary or invariant one. Both types do not take into account the data the good solution must represent, so convergence time is long. For problems like ours, studies (e.g., [Clark, 1976; Spence and Young, 1978]) suggest that these starting diagram types are more likely to direct the optimization to a local minimum. If the starting diagram is arbitrary as in example VennMaster [Kestler et al., 2008] the drawing method is non-deterministic as a different diagram is generated for the same area specification every time the optimization is run.

Both types of starting diagrams were considered for euler*APE*. The **arbitrary starting diagram** was a wellformed Venn diagram comprised of three ellipses whose properties were assigned random values. The **invariant starting diagram** was a wellformed Venn diagram with 3-fold rotational symmetry, comprised of three equally-sized circles and zones that were similar in size, except for those in only one curve that were around three times as much as the rest. The radii of the circles and the zone areas in the invariant starting diagram were predefined irrespective of the required quantitative data. As expected, both starting diagram types led to poor results in terms of generation time and diagram accuracy. The arbitrary type was particularly inappropriate, as our objective was to develop a deterministic method.

A **rational starting diagram** that is adapted to the required area specification is more effective, as it reduces convergence time and the likelihood of reaching a local minimum, without the need for a computationally expensive global search optimization (Section 5.6.1). So the optimization is more efficient. Thus, devising a method that generates an appropriate rational starting is worth the effort.

The circle-based method `venneuler` [Wilkinson, 2012] also uses a rational starting diagram that is close to the required solution. The required area of a curve in the diagram is the sum of the required area of each zone in that curve, so the radii of the circles for these curves can be computed analytically. To find appropriate centres for these circles, `venneuler` uses a distance matrix. This matrix only considers overlaps between every pair of ellipses, so the actual zone areas will not be close to the required, even if the overlapping area of each pair of ellipses is equal or close to the required (as shown in Section 3.7.6 subsection 'Constructing Accurate Area-proportional Venn Diagrams with Two But Not Three Circles').

Generating a rational starting diagram with ellipses is more difficult than circles. Given the required area of an ellipse, the two semi-axes of the ellipse cannot be computed analytically and an optimization technique is required to find appropriate values for these semi-axes (Section 3.7.6 subsection 'The Need for a Heuristic'). Besides the centre and the angle of rotation, the semi-axes of the ellipses in the starting diagram are also unlikely to be equal to those of the good solution. The optimization will then adjust all the properties of the ellipses in the starting diagram to search for a good solution.

We devised two methods to generate a rational starting diagram with ellipses, that is close to the required solution given an area specification and that has appropriate characteristics that could prevent convergence to a local minimum. We discuss these two methods in the following section, followed by the evaluation of both methods and the rationale of our chosen method for eulerAPE.

Our Novel Starting Diagram Generators

The starting diagram generated by both methods is drawn using **three ellipses with equal semi-axes**, so that the ellipses are depicted as circles, as given the area of an ellipse, its semi-axes cannot be computed analytically unless they are equal to one another. Also, area-proportional Venn diagrams drawn with circles are preferred (Section 3.7.1) and easier to comprehend (Section 3.5.2). So whenever possible, we wanted eulerAPE to draw the curves as circles or as ellipses with a circle-like shape (i.e., with similarly sized semi-axes). With equal semi-axes, any angle of rotation could be assigned to the ellipses, but an angle of rotation of 0, $\pi/3$ and $2\pi/3$ is assigned to the three respective ellipses, to ensure that the entire space of possible rotations is considered. The ellipses in the starting diagram need a centre. Assigning an appropriate centre is difficult and yet important as this determines the accuracy of the zone areas.

An accurate area-proportional Venn diagram can be drawn with circles for any data with two but not three sets (Section 3.7.6). The ellipses in our starting diagram are depicted as circles and so, an area-proportional **2-Venn diagram** can be drawn to accurately depict two of these overlapping ellipses using Chow and Ruskey's [2004] bisection method (explained in Section 3.7.6). A third ellipse must be added.

Only checking the accuracy of the overlapping area of every pair of ellipses (as in `venneuler`) does not warrant zone areas that are close to the required, and in most cases, all of the zone areas are far from the required (demonstrated in Section 3.7.6). Any changes carried out to any of the ellipses during the optimization will affect the area of **the zone that is located in exactly the three ellipses** and so, having a starting diagram that minimizes the error of this zone independently of the other zones in the diagram seems beneficial. This is so, as though it is important to have a starting diagram whose zone areas are all relatively close to the required solution, it is also important to have a starting diagram with appropriate characteristics that could prevent convergence to a local minimum. Maximizing the accuracy of the area of the zone located in exactly the three ellipses irrespective of the other zones could aid with the latter. When we discussed our evaluation and experimental comparison of our cost function with others in Section 5.6.4 and Appendix A, we gave an example in Figure A.8 that illustrates that sometimes, during the optimization, the area of the zone located in exactly the three ellipses is minimized, even though a larger area is required, up to the point that it is barely visible and a local minimum is reached. As demonstrated in Section 5.6.4, our specially devised cost function can identify and avoid such paths. However, starting with a diagram whose zone located in exactly the three ellipses already has the required area, further aids to avoid such local minima. As we shall demonstrate in subsection 'Evaluation of Our Proposed Generators', the overall *diagError* (Section 5.6.3) of the starting diagram

would still be reasonably low. This is primarily due to the choice of the ellipses for the generation of the accurate area-proportional 2-Venn diagram. Choosing the two ellipses that are either the ones with the largest required overlap or the largest two required ellipses in the diagram, increases the likelihood that the overall starting diagram is somewhat close to the required, as the accurate 2-Venn diagram would represent a large portion of the starting diagram and the third ellipse added to it is smaller than the other two, thus easier to position appropriately.

Hence, both **starting diagram generators**, S1 and S2, use the bisection method [Burden and Faires, 2010] to find a centre for the third ellipse (depicted as a circle) that minimizes the error of the area of the zone that is located in exactly the three ellipses independently of the other zones in the diagram. S1 and S2 mainly differ in the way they find the coordinates of the centre of the third ellipse:

- S1. Bisecting the interval along the *y*-axis and then along the *x*-axis;
- S2. Bisecting the interval along a bisector line.

The **general algorithm** for S1 and S2 to generate a rational starting diagram with respect to a given area specification ω is defined in Algorithm 5.2. S1 and S2 are discussed after Algorithm 5.2.

Algorithm 5.2: *GenerateARationaleStartingDiagram* (ω)

Input: ω, a set of seven quantities each corresponding to a zone of a 3-Venn diagram
Output: a rationale starting diagram with ellipses for ω

1. ω' ← the set of scaled quantities of ω (as defined in Section 5.6.2)
2. C ← the set of labels describing the required curves of the diagram to be generated for ω
3. **for** each curve label c in C **do**
4. $reqCurveAreas\,[c]$ ←$\sum \omega'[c], \omega'[c, c_1], \omega'[c, c_2], \omega'[c, c_1, c_2]$
                       where c_1 and c_2 are the two other curve labels in C
5. $radii\,[c]$ ← $\sqrt{(reqCurveAreas\,[c]\,/\,\pi)}$
6. **end for**
7. $rotations$ ← $\{0, \pi\,/\,3, 2\pi\,/\,3\}$
8. **if** S1 is used **then**
9.   **for** each pair of curve labels c_1 and c_2 in C **do**
10. $reqPairwiseOverlapAreas\,[c_1, c_2]$ ← $\sum \omega'[c_1, c_2], \omega'[c_1, c_2, c_3]$
                          where c_3 is the other curve label in C
11.   **end for**
12. C_{2Venn} ← the label of the curves with the largest required pairwise overlap area in
             reqPairwiseOverlapAreas
13. **else if** S2 is used **then**
14. C_{2Venn} ← the label of the curves with the largest two required curve areas in *requiredCurveAreas*
15. **end if**
16. ω'_{2Venn} ← the set of three quantities each corresponding to a zone of a Venn diagram with only the
             two curves in C_{2Venn}, computed using ω'
17. d_{2Venn} ← an accurate area-proportional 2-Venn diagram drawn with circles with respect to ω'_{2Venn},
             generated using Chow and Ruskey's [2004] bisection method
18. i_1 ← the upper intersection point (xi_1, yi_1) of the two circles in d_{2Venn}
19. i_2 ← the lower intersection point (xi_2, yi_2) of the two circles in d_{2Venn}

20. $c_3 \leftarrow$ the label of the curve in C but not in C_{2Venn}

21. $e_3 \leftarrow$ a circle with label c_3, radius $radii[c_3]$ and an undefined centre

22. $centres[c_3] \leftarrow$ the centre of e_3 computed by S1 or S2 using i_1 and i_2, such that when e_3 is
        added to d_{2Venn}, a wellformed 3-Venn diagram d_{3Venn} is formed and
        $\left| \omega'[c_1, c_2, c_3] - zas[c_1, c_2, c_3] \right| \rightarrow 0$, where c_1, c_2 are the curve labels in C that are not c_3
        and zas is the set of the actual zone areas in d_{3Venn}

23. **for** each curve label c in C_{2Venn} **do**

24. $centres [c] \leftarrow$ the centre of the circle with label c in d_{2Venn}

25. **end for**

26. **for** each curve label c in C **do**

27. $ellipses [c] \leftarrow$ an ellipse with label c, centre $centres[c]$, both semi-axes $radii[c]$ and angle of
        rotation $rotations[c]$

28. **end for**

29. **return** *ellipses*

The main objective of step 22 in Algorithm 5.2 is to find a centre for the third ellipse e_3, such that the area of the zone in exactly the three ellipses is closest as possible to the required scaled quantity irrespective of the other zones in the diagram. So the absolute error based on the scaled quantity in ω' (Section 5.6.2) is considered, rather than *zoneError* (Section 5.6.3) which takes into account the total diagram area. In eulerAPE, the zone areas are computed using our analytic method M2 in Section 5.4.3.

If **S1** is used, the centre of e_3 is obtained by applying: (1) the bisection method in an interval along the *y*-axis to obtain an appropriate *y*-coordinate, yc, while the *x*-coordinate is set to xi_1; (2) the bisection method in an interval along the *x*-axis to find an appropriate *x*-coordinate, xc, with respect to yc (as in Figure 5.17B), so that (xc, yc) is the centre of e_3. Let $r \in \mathbb{R}^+$ be the value assigned to the semi-axes of e_3. Bisection (1) is in the interval $(yi_1 + r, yi_1 - r)$ when $(yi_2 \leq (yi_1 - 2r))$ else in the interval $(yi_1 + r, yi_2 + r)$. Bisection (2) is in the interval $(xi_1 - d, xi_1 + d)$ where $d = \sqrt{(r^2 - (yi_1 - yc)^2)}$. The faded blue circles in Figure 5.17 indicate the placement of the third ellipse (depicted as a circle) for the endpoints of the interval in which the bisection method is applied and demonstrate that the endpoints of the interval as well as any value that is not in the interval will generate a diagram that is not a Venn diagram.
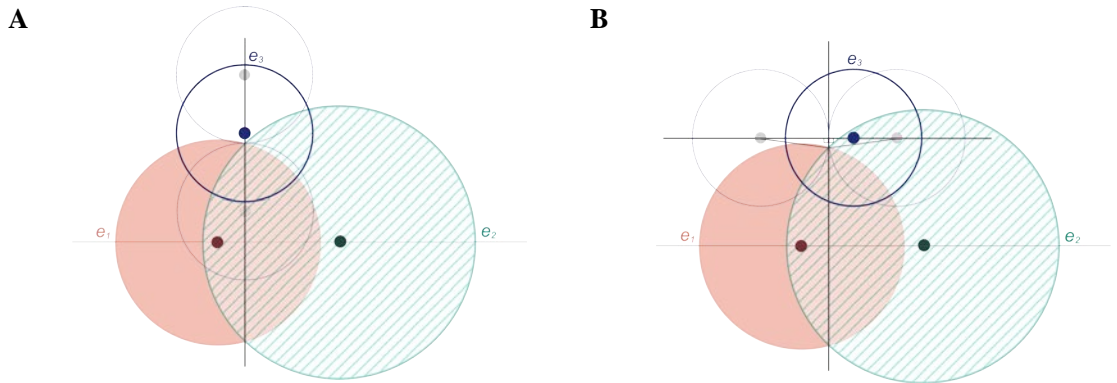
Figure 5.17: *Starting diagram generator S1—bisecting the interval along the y-axis and then along the x-axis.* The centre of ellipse e_3 is obtained by applying the bisection method in: (A) an interval along the *y*-axis to find the *y*-coordinate; (B) an interval along the *x*-axis to find the *x*-coordinate. The obtained centre should minimize the difference of the required and actual area of the zone in exactly the three ellipses.

If **S2** is used, the centre of e_3 is obtained by applying the bisection method in an interval along a line, *L*, as shown in Figure 5.18, where *L* is the bisector of the angle, ψ, between two lines, T_1 and T_2, that are respectively tangents to the other two ellipses e_1 and e_2 at the upper intersection point i_1. The interval along *L* is (*u*, *l*) where *u* is a point that lies above another point *l* on *L*, such that, as shown by the faded blue circles in Figure 5.18, the centre of e_3 must be between *u* and *l* (but not equal to *u* or *l*) for e_3 to intersect each of e_1 and e_2 twice, so that the seven zones interior to the curves of a 3-Venn diagram are formed. As illustrated by the faded blue circles in Figure 5.18, the endpoints of the interval as well as any value that is not in the interval will generate a diagram that is not a Venn diagram.
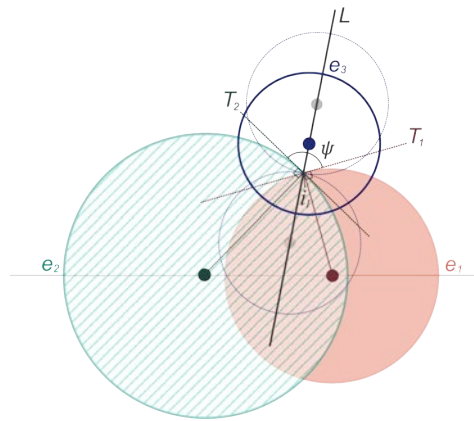
Figure 5.18: *Starting diagram generator S2—bisecting the interval along a bisector line.* The centre of ellipse e_3 is a point on the line *L* bisecting the angle ψ between the two tangents T_1 and T_2. The bisection method is applied in the interval indicated by the faded blue circles along *L*. The obtained centre should minimize the difference of the required and actual area of the zone in exactly the three ellipses.

As an example, Figure 5.19A and Figure 5.19B illustrate the starting diagram generated by respectively S1 and S2 for area specification ω = {*a*=1, *b*=10, *c*=1, *ab*=20, *ac*=1, *bc*=1, *abc*=10}. In both diagrams, the difference between the required scaled quantity and the area of the zone in exactly the three ellipses, which in this example is zone *abc*, is zero. The *diagError* for the diagram of S1 is 0.03 and for that of S2 is 0.02. Differences in the areas of the other zones are also noticeable. For instance, the areas of zones *ac* and *bc* are similar in the diagram of S2 in Figure 5.19B but different in that of S1 in Figure 5.19A. These differences in the starting diagram led the optimization (Section 5.6.6) to the different solutions in Figure 5.20, both of which have *diagError* zero with respect to ω.
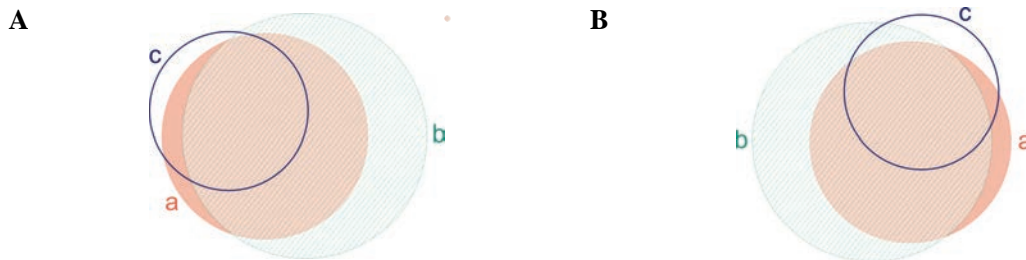
Figure 5.19: *The starting diagrams generated by S1 and S2 for the same area specification.* For area specification ω = {*a*=1, *b*=10, *c*=1, *ab*=10, *ac*=1, *bc*=1, *abc*=10}, S1 generated A and S2 generated B.
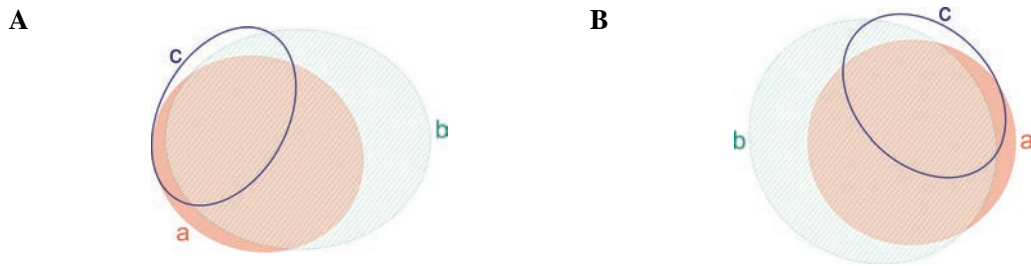
Figure 5.20: *The diagrams generated by our hill climber for the area specification and the starting diagrams in Figure 5.19.* The area specification is $\omega = \{a=1, b=10, c=1, ab=10, ac=1, bc=1, abc=10\}$. Both diagrams have *diagError* zero with respect to ω. Our hill climber generates A if Figure 5.19A (generated by S1) is used as a starting diagram and B if Figure 5.19B (generated by S2) is used as a starting diagram.

Evaluation of Our Proposed Generators

Figure 5.21 shows the *diagError* Equation (5.18) of the starting diagrams generated by S1 and S2 for the two sets of 10,000 area specifications obtained from DIAGLIB_*design* and AREASPECLIB_*design* (Section 5.5), together with the time taken to generate the diagrams on an Intel Core i7-3770 CPU @3.4GHz with 8GB RAM, 64-bit Microsoft Windows 7 Professional SP1 and Java Platform 1.7.0_10. In Figure 5.21 and this section, the starting diagrams generated for the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design* are respectively referred to as **DL***sd* and **ASL***sd*.

The value of *diagError* is always in [0,1] (Section 5.6.3) and so, **the errors of DL***sd* **and ASL***sd* **for both S1 and S2 are relatively low**. The mean *diagError* of DL*sd* and ASL*sd* are respectively 0.04 and 0.11 for S1 and 0.05 and 0.12 for S2. The mean error of ASL*sd* is greater than that of DL*sd*, as in contrast to the latter, an accurate area-proportional 3-Venn diagram drawn with ellipses might not exist for some of the area specifications in the former. Also, 76% and 18% of respectively DL*sd* and ASL*sd* for S1 and 63% and 15% of respectively DL*sd* and ASL*sd* for S2 had a *diagError* $\leq$ 0.05 (i.e., 5%). Looking at the time required to generate the starting diagram, S1 is overall marginally faster than S2, with a mean of 3 milliseconds for S1 and a mean of 8 milliseconds for S2. However, the response time for both is more than 10 times less than the 0.1 second limit for an instantaneous response [Card et al., 1991; Miller, 1968]. So, in terms of generation time, any one of the generators could be used as neither of them would affect the overall performance of our drawing method.
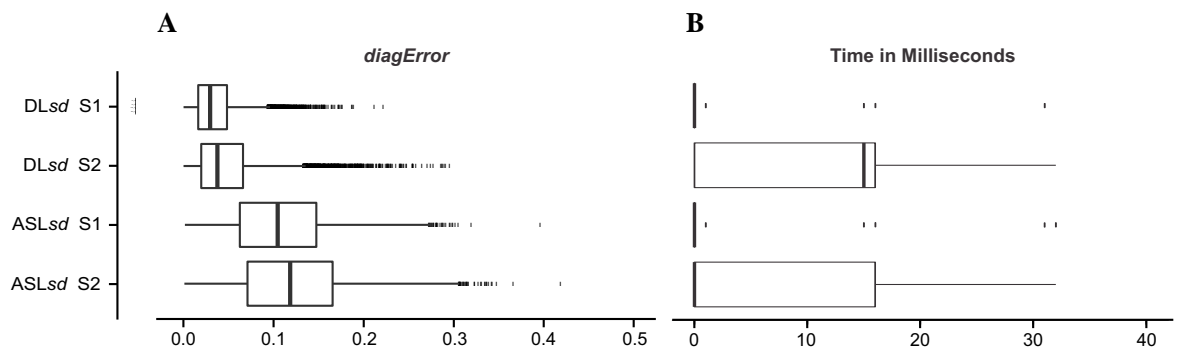
Figure 5.21: *The diagError of the starting diagrams generated by S1 and S2 for the area specifications obtained from libraries DIAGLIB_design and AREASPECLIB_design and the time taken in milliseconds to generate the diagrams.* In these plots, the starting diagrams generated for the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design* are referred to as DL*sd* and ASL*sd* respectively. DL*sd* and ASL*sd* each have 10,000 starting diagrams, one for each of the 10,000 area specifications.

These results indicate that both **S1 and S2 generate starting diagrams that are close to the required solution and in relatively fast time**, with the majority having a *diagError* percentage of less than 5% whenever a good solution to the required area specification is known to exist. S1 seems to generate diagrams with a *diagError* that is marginally smaller than that of S2. However, as we shall demonstrate, the diagrams generated by S1 might not fully satisfy our objective, that of finding an appropriate centre for the third ellipse that minimizes the error of the zone located in exactly the three ellipses irrespective of the other zones.

As explained earlier, S1 obtains the centre (*xc*, *yc*) of the third ellipse by first applying the bisection method in an interval along the *y*-axis to find *yc* while *xc* is set to the *x*-coordinate, xi_1, of the upper intersection of the two ellipses in the 2-Venn diagram and then, by applying the bisection method in an interval along the *x*-axis to find an appropriate *xc* with respect to *yc*. However, often, no interval that ensures that the diagram is a 3-Venn diagram exists in which the bisection method can be applied to find an appropriate *xc* and so, *xc* is set to xi_1. This occurred during the generation of 40% and 43% of respectively DL*sd* and ASL*sd* (this could have contributed to the faster generation time of S1 with respect to S2, as in these cases, the bisection method was applied on only one of the coordinates of the centre). In such cases, the obtained centre might not be the most appropriate, and for some diagrams, the area of the zone located in exactly the three ellipses will not be equal to the required area. In fact, the area of this zone in DL*sd* and ASL*sd* (in total 20,000 diagrams) was on average equal to that required for those generated by S2, but on average 5% greater or smaller than that required for those generated by S1. Figure 5.22A illustrates an example of such a starting diagram generated by S1 for $\omega = \{a{=}1, b{=}20, c{=}200, ab{=}1, ac{=}1, bc{=}200, abc{=}200\}$. Its *diagError* is 0.06, yet zone *abc* is 20% smaller than required. In contrast, the starting diagram in Figure 5.22B generated by S2 for ω, has a zero *diagError* and the area of zone *abc* is as required. So a better compromise can be achieved by obtaining the centre from a single application of the bisection method along the *x*-axis and the *y*-axis simultaneously (as in S2), rather than obtaining the individual coordinates from two separate applications of the bisection method (as in S1). Thus, **S2 is more effective than S1**.
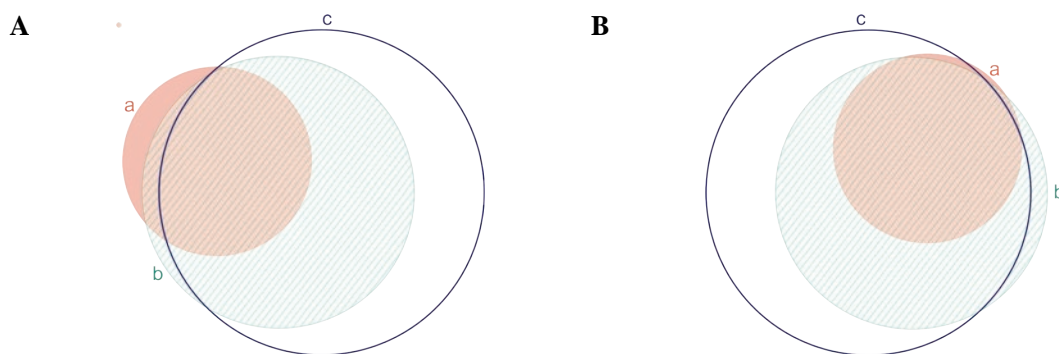
Figure 5.22: *Illustrating a limitation of S1 in contrast to S2.* A and B are the starting diagrams generated by S1 and S2 respectively for the area specification $\omega = \{a{=}1, b{=}20, c{=}200, ab{=}1, ac{=}1, bc{=}200, abc{=}200\}$. The area of zone *abc* in B is equal to the required area, but that in A is not. In this example, S1 could not apply the bisection method along the *x*-axis.

Also, in a preliminary experimentation using our optimization algorithm (Section 5.6.6) over two sets of 10,000 random area specifications obtained from a library of type DIAGLIB and another of type AREASPECLIB (Section 5.5.1), **S2 performed better than S1** by 0.5% in generating a good solution for the area specifications obtained from the DIAGLIB type library, while there was no difference between

S1 and S2 in generating a good solution for the area specification obtained from the AREASPECLIB type library. A good solution exists for the area specifications obtained from DIAGLIB (as such libraries are a collection of wellformed 3-Venn diagrams drawn with ellipses with random properties) and so, failing on such area specification means that a local minimum is encountered. Thus, S2 should be preferred.

The Generator Employed By eulerAPE

As discussed in the previous section, the starting diagram generator S2 is more effective than S1 and is thus employed by eulerAPE.

As we demonstrate in Section 5.7, using a rational starting diagram generator (S2), eulerAPE is more likely to draw a diagram within the 10 second response time limit that ensures users' attention is maintained [Card et al., 1991; Miller, 1968], is less likely to converge to a local minimum (Section 5.7.1) and is more likely to generate diagrams with circle-like curves (i.e., ellipses whose individual semi-axes are similar in size) whenever possible. In some cases, the starting diagram is so close to the required solution, that the user might consider it good enough and does not run the optimization (e.g., [Palagi et al., 2013]). Above all, eulerAPE is deterministic.

We will now discuss our optimization algorithm that uses such a rational starting diagram to generate the required good solution with respect to a given area specification.

5.6.6 The Optimization Algorithm

As discussed and justified in Section 5.6.1, we adopted a simple hill-climbing algorithm that commences with a rational starting diagram and systematically adjusts the properties of its ellipses that minimize our cost function (Equation (5.20)), until a good diagram (Equation (5.19)) with respect to the given area specification is obtained. Though simple and a local search, it rarely encounters a local minimum and if it does, our algorithm is capable of handling such cases (Section 5.6.7; also Section 5.6.1) and obtain a good solution whenever an accurate area-proportional 3-Venn diagram drawn with ellipses is known to exist for the given area specification (Section 5.7.1).

Our optimization algorithm is characterized by the following three parameters that determine how, at every iteration, each ellipse e is modified in search for other possible solutions:

- $p\gamma$, the number of pixels by which one or both coordinates of the centre of e are increased or decreased to explore eight new centres for e, as illustrated in Figure 5.23A, where the black ellipse and the black point are respectively e and its centre prior to any change and the grey points are the eight new centres for e;

- $p\alpha\beta$, the scaling percentage by which one or both semi-axes of e are increased or decreased to explore eight new semi-axes for e, as illustrated in Figure 5.24A-D, where the solid black ellipse is e prior to any change and the dashed coloured ellipses are e with the eight new semi-axes;

- $p\theta$, the number of radians by which the angle of rotation of e is increased or decreased to explore two new angles of rotation for e, as illustrated in Figure 5.23B, where the solid black ellipse is e prior to any change and the dashed coloured ellipses are e with the two new angles of rotation.

New properties that are explored for each ellipse are accepted only if the diagram cost is improved (i.e, reduced). This cost is computed using Equation (5.20) and its sole purpose is to determine whether any of the above new ellipses' properties should be accepted and so, drive the optimization to the good solution. The diagram goodness with respect to a given area specification is determined by our goodness measure in Equation (5.19), once the new properties are explored (possibly accepted) for all the ellipses.
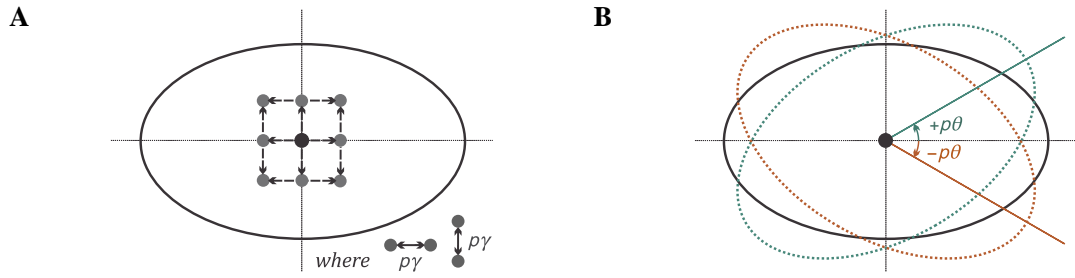
Figure 5.23: *The different ways how the centre and the angle of rotation of every ellipse at every iteration are respectively modified using the parameters $p\gamma$ and $p\theta$ in search for other solutions.* (A) The grey points indicate the new centres that are obtained when the one or both of the coordinates of the centre of an ellipse (solid black) are increasing or decreasing by $p\gamma$. (B) The dashed ellipses indicate how an ellipse (solid black) is changed when its angle of rotation is increased or decreased by $p\theta$.
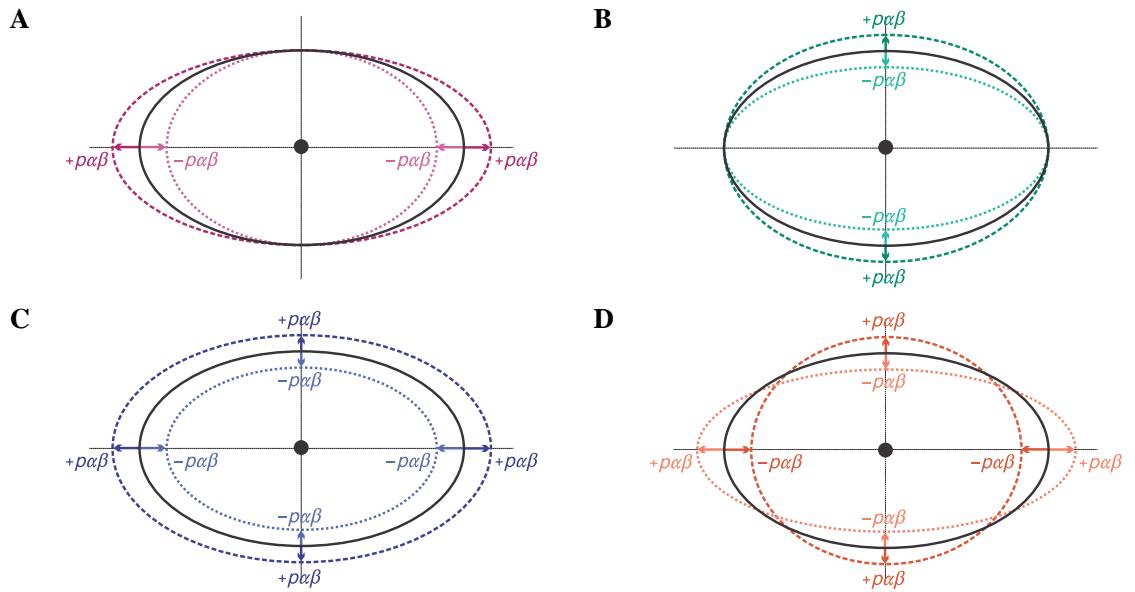
Figure 5.24: *The different ways how the semi-axes of every ellipse at every iteration is modified using parameter $p\alpha\beta$ in search for other solutions.* A label $+p\alpha\beta$ means that semi-axis was increased by the percentage in $p\alpha\beta$, while $-p\alpha\beta$ means that semi-axis was decreased by the percentage in $p\alpha\beta$. The dashed ellipses indicate how an ellipse (solid black) is changed when: (A) only the semi-major axis is increased or decreased by $p\alpha\beta$; (B) only the semi-minor is increased or decreased by $p\alpha\beta$; (C) the semi-axes are both increased or both decreased by $p\alpha\beta$; (D) one of the semi-axes is increased and the other is decreased by $p\alpha\beta$.

At the start, $p\gamma = 10$ pixels, $p\alpha\beta = 5\%$ and $p\theta = 2\pi/3$. For every iteration with no changes to a property of the ellipses, the value of the respective parameter is reduced linearly (halved), so that major changes only occur at the start when the search space is explored for a good yet non-refined solution. In later iterations, when the values of $p\gamma$, $p\alpha\beta$ and $p\theta$ are reduced, minor changes to the diagram are

explored, so that the diagram is refined to the required good solution. This acts as a cooling schedule similar to that used in the global optimization method of simulated annealing. If such a cooling schedule is not used, $p\gamma$, $p\alpha\beta$ and $p\theta$ would be invariant and their values would have to be significantly small for any changes to improve the diagram. This would lead to longer convergence time as changes are minor and numerous changes are required for the good solution to be reached. So, if the values of $p\gamma$, $p\alpha\beta$ and $p\theta$ are not small enough or if it is too computationally expensive for all the required changes to be carried out, the optimization would converge to a local minimum. Thus, by allowing the exploration of major changes to the diagram at the start of the optimization, our cooling schedule can reduce the likelihood of converging to a local minimum and the time required to converge to the good solution.

The starting values for $p\gamma$, $p\alpha\beta$ and $p\theta$ were chosen after the diagram generation process for different area specifications was observed using different starting values for $p\gamma$, $p\alpha\beta$ and $p\theta$. The chosen starting values were the most effective for most of the area specifications. We wanted these values to be as large as possible to explore major initial changes for a more adequate exploration of the search space. However, we noted that very large values often lead to diagrams that are not Venn diagrams and so, improved solutions are only obtained after several iterations. We did however observe that some cases can benefit from starting values that are 20% larger than those we have selected. So, when eulerAPE encounters a local minimum, if enabled, it reruns the optimization using starting values for $p\gamma$, $p\alpha\beta$ and $p\theta$ that are 20% larger than those used in the previous run, to further explore the search space. As we demonstrate in our evaluation in Section 5.7.1, this technique is effective in handling the very few cases when eulerAPE encounters a local minimum.

The cooling rate (i.e., to half the values of $p\gamma$, $p\alpha\beta$ and $p\theta$) was chosen in a similar manner after informal experimentation using different rates. We observed that faster cooling rates can limit the number of refinements carried out on the diagram such that the required good solution is not achieved, while slower cooling rates are more computation expensive.

Hence, **the general algorithm** to generate an accurate wellformed area-proportional 3-Venn diagram with respect to a given area specification ω, using ellipses, is as follows (Algorithm 5.3):

Algorithm 5.3: eulerAPE (ω)

Input: ω, a set of seven quantities each corresponding to a zone of a 3-Venn diagram
Output: an area-proportional 3-Venn diagram drawn with ellipses for ω and whether the diagram is
   accurate or inaccurate

1. $d \leftarrow$ a rational starting diagram with respect to ω
2. **if** d is a good diagram for ω by Equation (5.19) **then**
3.   **return** d, *accurate*
4. **end if**
5. $\omega' \leftarrow$ the set of scaled quantities of ω (as defined in Section 5.6.2)
6. $p\gamma \leftarrow 10$ pixels, $p\alpha\beta \leftarrow 5\%$, $p\theta \leftarrow 2\pi / 3$
7. **do**
8.   **for** each ellipse e in d **do**
9.   *centres* $\leftarrow$ the eight centres obtained by $p\gamma$ for e
10.   **for** each c in *centres* **do**
11.   **if** the cost of d by Equation (5.20) is reduced when the centre of e in d is c **then**

12.   Change the centre of *e* in *d* to *c*
13.   **end if**
14.   **end for**
15.   *semiaxes* ← the eight semi-axes obtained by *pαβ* for *e*
16.   **for** each *s* in *semiaxes* **do**
17.   **if** the cost of *d* by Equation (5.20) is reduced when the semi-axes of *e* in *d* are *s* **then**
18.   Change the semi-axes of *e* in *d* to *s*
19.   **end if**
20.   **end for**
21.   *rotations* ← the two angles of rotation obtained by *pθ* for *e*
22.   **for** each *r* in *rotations* **do**
23.   **if** the cost of *d* by Equation (5.20) is reduced when the angle of rotation of *e* in *d* is *r* **then**
24.   Change the angle of rotation of *e* in *d* to *r*
25.   **end if**
26.   **end for**
27.   **end for**
28.   **if** no ellipse in *d* had its centre changed **then**
29.   Divide *pγ* by 2
30.   **end if**
31.   **if** no ellipse in *d* had any of its semi-axes changed **then**
32.   Divide *pαβ* by 2
33.   **end if**
34.   **if** no ellipse in *d* had its angle of rotation changed **then**
35.   Divide *pθ* by 2
36.   **end if**
37.   **if** $p\gamma \leq \varepsilon$, $p\alpha\beta \leq \varepsilon$ and $p\theta \leq \varepsilon$, where $\varepsilon = 10^{-6}$ **then**
38.   **return** *d*, *inaccurate*
39.   **else**
40.   **if** *d* is a good diagram for *ω* by Equation (5.19) **then**
41.   **return** *d*, *accurate*
42.   **end if**
43.   **end if**
44. **loop**

As mentioned earlier, when a local minimum is reached, euler*APE* **can rerun the optimization** (if the rerun option is enabled) using starting values for *pγ*, *pαβ* and *pθ* that are 20% larger than those used in the previous run. So, if the optimization terminates at step 38 of Algorithm 5.3, *pγ*, *pαβ* and *pθ* are set back to the starting values as in step 6 and each value is incremented by 20%. The optimization is then run again starting from step 7. If it terminates again at step 38, *pγ*, *pαβ* and *pθ* are set back to the starting values of that run and incremented by another 20%. So, if $m \in \mathbb{R}^+$ is the percentage by which the starting values of the parameters are iteratively incremented for every rerun (i.e., 20% in euler*APE*) and *j* $\in \mathbb{N}$ is the number of times the optimization was previously run, each of *pγ* = 10 pixels, *pαβ* = 5%, *pθ* = 2π/3 are multiplied by $(1 + m)^j$ before the next rerun commences. Whenever this rerun option is enabled, euler*APE* terminates either when a good diagram is found or when 10 reruns are completed and

a good diagram is yet not found, in which case the diagram with the lowest *diagError* (Equation (5.18)) out of the 11 diagrams each of which were generated per run is returned.

At every iteration of our optimization algorithm, any of the properties of any of the ellipses can be changed and similarly, the value of any of the parameters can be reduced independently of each other. However, **our algorithm could have been designed** so that only one change per iteration is allowed or our cooling schedule reduces the value of the parameters simultaneously. We discuss these alternatives and our preliminary experimentation that led us to our current design in the next subsection. We have also tried to reduce the dimensionality of our optimization problem by not changing any of the properties of the largest ellipse in the starting diagram. However, as we discuss in one of the next subsections, such a restriction could lead the optimization to converge on the wrong solution.

Since this work focuses on wellformed 3-Venn diagrams, the diagrams always have three curves and seven zones interior to the curves (each zone always has one minimal region, as the diagrams are wellformed) and so, the number of steps per iteration of the optimization algorithm (i.e., the number of steps in 7 of the algorithm) is constant. Thus, the time complexity of the general algorithm depends on the maximum number of iterations that are required for the optimization to converge to a solution (i.e., the number of times 5 of the algorithm is iterated through), but this cannot be determined for the generation of an area-proportional Venn diagram with respect to any possible area specification.

Choosing a Design—Single or Multiple Changes, Dependent or Independent Parameters

The number of changes (to the centre, semi-axes, angle of rotation of the ellipses) allowed per iteration and the way the parameters (i.e., $p\gamma$, $pα\beta$, $p\theta$) are reduced could affect the quality of the generated diagrams and the convergence time. Finding the most effective configuration for each and for the combination of the two is important. Hence, when devising our algorithm, the following four possible designs were identified and evaluated:

- SC–DP.   SINGLE CHANGE–DEPENDENT PARAMETERS;
- SC–IP.   SINGLE CHANGE–INDEPENDENT PARAMETERS;
- MC–DP. MULTIPLE CHANGES–DEPENDENT PARAMETERS;
- MC–IP.   MULTIPLE CHANGES–INDEPENDENT PARAMETERS.

SINGLE CHANGE (SC) is when at every iteration at most a single property of a single ellipse is changed, where the single change that reduces the cost of the diagram most is chosen.

MULTIPLE CHANGES (MC) is when at every iteration any of the properties of any of the ellipses could be changed to one that reduced cost of the diagram most.

DEPENDENT PARAMETERS (DP) is when $p\gamma$, $pα\beta$ and $p\theta$ are reduced (in our case, divided by 2) simultaneously when none of the properties of any of the ellipses have been changed during that iteration. So, for instance, if the centre and the semi-axes were not changed for any of the ellipses, but the angle of rotation was changed for one or more ellipses, then none of the parameters are reduced as they are dependent on each other.

INDEPENDENT PARAMETERS (IP) is when $p\gamma$, $pα\beta$ and $p\theta$ are reduced (in our case, divided by 2) independently of each other when the property they handle was not changed for any of the ellipses during that iteration. So example, if the centre was not changed for any of the ellipses, the value of $p\gamma$ is reduced, irrespective of whether the semi-axes or angle of rotation of any ellipse was changed.

In a preliminary experimentation, we evaluated these four designs over a set of 5,000 random area specifications obtained from a library of type DIAGLIB (Section 5.5.1). Such a library is a collection of wellformed 3-Venn diagrams drawn with ellipses with random properties. So, a solution is known to exist for each of the evaluated area specifications and failing to produce a good diagram for any of these area specifications means that a local minimum is reached.

MC–IP performed best in generating a good diagram for the given area specifications, followed by MC–DP, SC–DP and SC–IP (MC–IP performed better than MC–DP, SC–DP and SC–IP by respectively 1.9%, 3.1% and 26.3%). MC–IP was also the fastest in converging to a solution, once again followed by MC–DP, SC–DP and SC–IP (MC–IP was 2.3, 2.8 and 11.6 times faster than respectively MC–DP, SC–DP and SC–IP). A longer convergence time was expected for those carrying out a single change per iteration (SC), yet here we also note that this increases the likelihood of converging to a local minimum and it is thus more effective to allow multiple changes per iteration (MC). Comparing the results for MC–DP and MC–IP, we note that having independent parameters is more effective in obtaining a good diagram with respect to the given area specification and in convergence time.

Thus, we adopted MC–IP for the eulerAPE's optimization algorithm, which we have discussed earlier.

Trying to Reduce the Dimensionality of the Optimization Problem

Our optimization algorithm allows changes to all the properties (i.e., the centre, semi-axes and angle of rotation) of all the ellipses. The optimization problem can be reduced by not changing any of the properties of one the ellipses in the starting diagram, such that this ellipse is depicted as a circle.

We evaluated this possibility over the two sets of 10,000 area specifications obtained from the libraries DIAGLIB_*design* and AREASPECLIB_*design* (Section 5.5) by not allowing any changes to any of the properties of the largest ellipse in the starting diagram and the following results were obtained. In this case, the rerun option of the optimization algorithm was disabled. An accurate, good diagram was generated for 98.9% and 75.7% of respectively the area specifications obtained from DIAGLIB_*design* and those obtained from AREASPECLIB_*design*, after an average number of 65 iterations and 204 iterations for the respective sets of area specifications. However, when changes to any of the ellipses was allowed, an accurate, good diagram was obtained for 99.5% and 83.5% of respectively the area specifications obtained from DIAGLIB_*design* and those obtained from AREASPECLIB_*design*, after an average number of 73 iterations and 107 iterations for the respective sets of area specifications (these results were also discussed in Section 5.6.4). Thus, not changing one of the ellipses could minimize the dimensionality of the optimization problem, but it will increase the likelihood of converging to a local minimum and in general, increase the required number of iterations to get to a good solution.

Figure 5.25 shows an example where the optimization fails to generate a good, accurate diagram when one of the ellipses is not changed. An accurate area-proportional 3-Venn diagram had to be drawn with respect to the area specification $\omega = \{a=32086, b=16009, c=20393, ab=394, ac=740, bc=15776, abc=1\}$ obtained from DIAGLIB_*design*. When the largest ellipse (i.e., ellipse c) was not changed, the inaccurate diagram in Figure 5.25A with *diagError* = 0.03 was generated, but when any changes to any of the ellipses was allowed (as in eulerAPE), the accurate, good diagram in Figure 5.25B was obtained. While it was possible for both ellipses b and c in Figure 5.25B to be elongated to accurately depict the required zone areas, the circular shape of ellipse c in Figure 5.25A could not, so ellipse b was greatly elongated in an attempt to accurately depict the required zone areas. Yet at one point, ellipse b could not be stretched further (without breaking the diagram's wellformedness), so a local minimum was reached.
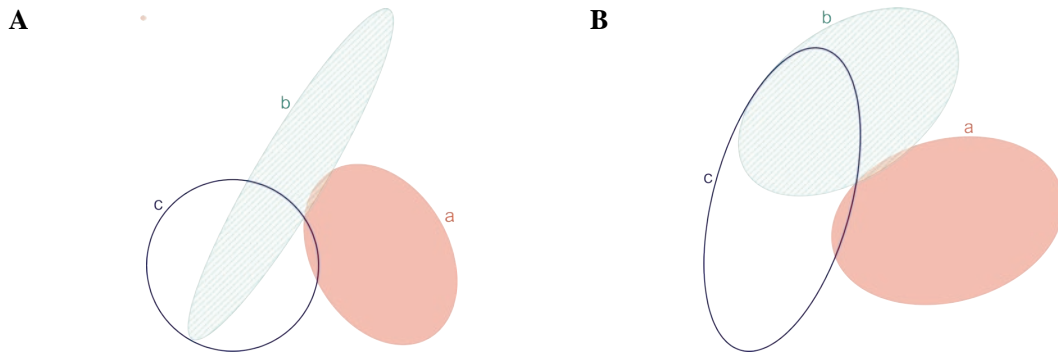
Figure 5.25: *An example where the optimization fails to converge to a good diagram when none of the properties of the largest ellipse in the starting diagram are changed.* The two diagrams were generated for area specification ω = {a=32086, b=16009, c=20393, ab=394, ac=740, bc=15776, abc=1}, which was obtained from DIAGLIB_*design*. (A) An inaccurate area-proportional 3-Venn diagram with respect to ω (*diagError* = 0.03) generated when the none of the properties of the largest ellipse in the starting diagram, in this case ellipse c, were changed. (B). A good, accurate area-proportional 3-Venn diagram with respect to ω generated with our optimization algorithm, which allows any properties of any of the ellipses to be changed.

As we have mentioned throughout our discussion on the design of the various of components that make up eulerAPE, a number of techniques have been adopted to avoid and handle local minima. So, despite adopting a simple local optimization hill-climbing algorithm, eulerAPE rarely reaches a local minimum and when it does, it is capable of handling it and yet, find a good solution when one is know to exist (Section 5.7.1). We summarize these techniques in the next section.

5.6.7 Avoiding and Handling Local Minima

As discussed in Section 5.6.1 and as demonstrated in eulerAPE's evaluation in Section 5.7, if techniques are adopted to avoid and handle local minima, a local search optimization like eulerAPE's simple hill climbing algorithm can be used to draw accurate area-proportional Venn diagrams in relatively fast time, without the need for complex and computationally expensive global optimization techniques.

For eulerAPE, we have adopted the following techniques to avoid and handle local minima:

- A rational starting diagram (Section 5.6.5)

    that is close to the required solution and has characteristics that could avoid local minima (e.g., the largest two ellipses make up an accurate area-proportional 2-Venn diagram with respect to the quantities these ellipses and their overlap depict; the error of the zone located in exactly the three ellipses is minimized with respect to the quantity it should depict and irrespectively of the other zones, as any changes carried out to the diagram during the optimization process will affect this zone);

- A cost function (Section 5.6.4)

    that prevents the optimization from taking paths that would improve the overall diagram at the expense of constantly diminishing the area of one zone that is meant to be larger, up to the point when this zone becomes close to non-existent and a local minimum is reached;

- A cooling schedule (Section 5.6.6)

  that reduces the value of the parameters of the optimization algorithm (i.e., $p\gamma$, $p\alpha\beta$ and $p\theta$ that determine how the ellipses are modified per iteration in search for other solutions) linearly, allowing these parameters to be initially large, so major changes to the diagram are done at the start when the search space is explored for a good yet non-refined solution, while minor changes are done later when the diagram is refined to get the required one;

- A rerun option (Section 5.6.6)

  that, if enabled, when a local minimum is encountered, the optimization is rerun using starting values for the parameters of the optimization algorithm (i.e., $p\gamma$, $p\alpha\beta$ and $p\theta$) that are 20% greater than those used in the previous run (to further explore the search space), until a good diagram is found or until 10 reruns are completed.

All of these techniques have been discussed throughout in Section 5.6 on the design of eulerAPE.

Other techniques, such as the tabu search (to ensure new paths in the search space are explored) and the steepest-descent method (to aid handling ridges), could have been adopted. Others like the random-restart hill climbing or the stochastic hill climbing were not appropriate as we wanted a deterministic method, and as discussed in Section 5.6.5, a random start can increase the likelihood of reaching a local minimum. Others like global optimization methods (e.g., simulated annealing, genetic algorithms, evolutionary algorithms, swarm algorithms) are complex and computationally expensive, and as discussed in Section 5.6.1, such methods are not necessary for our problem and not appropriate for obtaining a good solution in relatively fast time.

In our evaluation of eulerAPE (Section 5.7.1), eulerAPE rarely encountered a local minimum and when it did, a good diagram was still obtained when one was known to exist for the given area specification. This demonstrates that the adopted techniques are effective in both avoiding and handling local minima and with their use, a local search optimization such as the simple hill climbing algorithm can be used to effectively generate the required accurate area-proportional Venn diagram within an average time that is fast enough (~10 seconds; Section 5.7) to ensure users' attention is maintained [Card et al., 1991].

We will now discuss our evaluation of eulerAPE and the effectiveness of ellipses in drawing accurate area-proportional 3-Venn diagrams for various random area specifications.

5.7 The Effectiveness of eulerAPE and Ellipses

To evaluate the effectiveness of ellipses in drawing an accurate area-proportional 3-Venn diagram for a given area specification, we first evaluated the effectiveness of eulerAPE in drawing good diagrams for drawable area specifications, that is area specifications for which a good diagram is known to exist (Section 5.7.1). Being able to handle such area specifications means that eulerAPE can avoid and handle local minima and if eulerAPE cannot draw an accurate, good diagram for any of the area specifications in our second evaluation (Section 5.7.2) that were made up of random values, then it is highly likely that a good diagram drawn with ellipses does not exist for that area specification. In this way, we were able to identify characteristics of area specifications that are drawable or not with ellipses (Section 5.7.2). In our second evaluation, we also generated diagrams for the same random area specifications using a variant of eulerAPE that restricts the ellipses to circles to identify whether diagrams for any of the area specifications could have been drawn with circles(Section 5.7.2). We then compared these results to the

diagrams generated for the same random area specifications by venneuler [Wilkinson, 2012], which is considered to be the most effective circle-based drawing method (Section 5.7.3). Finally, we compared the accuracy and the curve aesthetics of the diagrams generated by eulerAPE and various other drawing methods using circles or polygons for real data in an application area (Section 5.7.4).

The error of the diagrams generated by eulerAPE and other drawing methods was measured using our *diagError* Equation (5.18) in Section 5.6.3, whose value is in [0,1]. **Good diagrams** are those which satisfy the diagram goodness measure Equation (5.19) in Section 5.6.3 and are thus wellformed 3-Venn diagrams with *diagError* $\leq 10^{-6}$. Wellformed 3-Venn diagrams are always generated by eulerAPE but not always by other methods. In our quantitative experiments, the number of iterations and the time taken to generate the diagrams was also recorded. All the experiments were run on an Intel Core i7-3770 CPU @3.4GHz with 8GB RAM, 64-bit Microsoft Windows 7 Professional SP1 and Java Platform 1.7.0_10.

As with the rest of this work, this evaluation focuses on area specifications corresponding to 3-Venn diagrams (reasons further discussed in Section 5.2.3). Some Euler diagrams can already be drawn with eulerAPE, but this needs to be evaluated further in the future.

5.7.1 For Drawable Area Specifications

To evaluate whether eulerAPE can generate a good diagram for area specifications for which such a diagram drawn with ellipses is known to exist (i.e., drawable area specifications) and whether it can handle and avoid local minima, we used eulerAPE to generate diagrams for the 10,000 area specifications obtained from DIAGLIB_*eval*. DIAGLIB_*eval* is a library of 10,000 wellformed 3-Venn diagrams drawn with ellipses labelled *a*, *b*, and *c* whose properties are assigned random values. The zone areas of each diagram comprise one of the 10,000 area specifications over which this experiment was conducted. Details about the library and the characteristics of the area specifications are discussed in Section 5.5.

The rerun option of our optimization algorithm (Section 5.6.6) was enabled to verify whether eulerAPE can still draw a good diagram if a local minimum is reached in the first run. While choosing an appropriate cost function for eulerAPE in Section 5.6.4, an experimental evaluation over a different set of area specifications of the same type using the same cost function and algorithm as that of eulerAPE but with the rerun option disabled, revealed that for 0.5% of the area specifications a non-good diagram was generated. So, for a few area specifications, eulerAPE might not generate a good diagram during the first run. If so, we need to investigate whether rerunning the optimization algorithm with different starting values for its parameters ($p\gamma$, $p\alpha\beta$ and $p\theta$) as discussed in Section 5.6.6 is effective in finding a good diagram, and how these reruns affect the performance of eulerAPE, in terms of time and iterations. The maximum number of allowed reruns is 10, after which, unless a good diagram is found, the diagram with the lowest *diagError* out of the 11 diagrams, each generated during one of the runs, is returned.

Results

By the first run, **good diagrams** were generated for 9939 of the 10,000 area specifications (i.e., 99.4%). Though for the remaining 61 area specifications (i.e., 0.6%) a non-good diagram was generated, the *diagError* of these diagrams was relatively low (median 1.06×10^{-4}, mean 2.38×10^{-3}, min 1.02×10^{-6}, max 3.09×10^{-2}), with 54 of them (i.e., 88.5%) having *diagError* ≤ 0.01. For all of these 61 area specifications, good diagrams were generated after rerunning the optimization; for the majority (38/61=62.3%), a good diagram was drawn after first rerun (Figure 5.26).
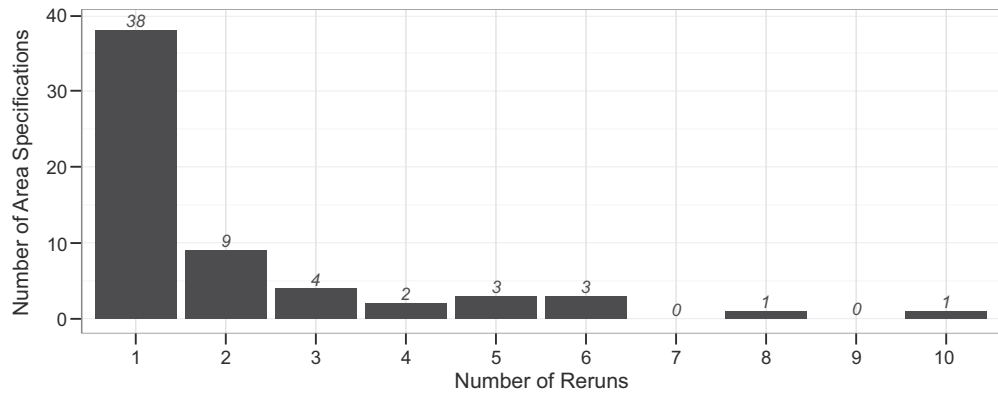
Figure 5.26: *The number of reruns (1-10) that were required for* eulerAPE *to generate a good diagram for the 61 area specifications obtained from* DIAGLIB_eval *for which a non-good diagram was generated during the first run.* Number of reruns: median 1, mean 2.1.

When the optimization algorithm is rerun, more **time and total number of iterations** are required to generate a good diagram (Figure 5.27). Even so, the generation of the 10,000 good diagrams had an overall median and mean time of respectively 0.4 seconds and 2.5 seconds, and an overall median and mean number of iterations of respectively 32 and 273. Also, for 97.7% of the 10,000 area specifications, a good diagram was generated within 1 second (98.1% of the 9939 good diagrams generated in run 1; 34.4% of the 61 good diagrams generated in a rerun) and for 99.7% of the 10,000 area specifications a good diagram was generated within 10 seconds (99.9% of the 9939 good diagrams generated in run 1; 62.3% of the 61 good diagrams generated in a rerun). These results are important as a response time of around 1 second ensures the users' train of thought is uninterrupted and a response time of around 10 seconds ensures the users' attention is retained [Card et al., 1991; Miller, 1968].
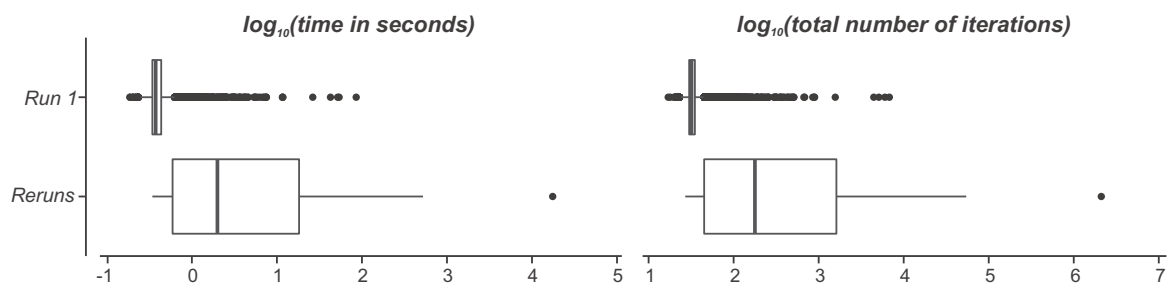
Figure 5.27: *The log_{10}(time in seconds) and log_{10}(total number of iterations) taken to generate good diagrams for 9939 of the area specifications in the first run (Run 1) and for 61 of the area specifications in any of the one to a maximum of 10 reruns (Reruns)—the 10,000 area specifications were obtained from* DIAGLIB_eval. *For Run 1 and Reruns*: time, medians 0.4 seconds and 2.0 seconds, means, 0.5 seconds and 329.3 seconds; number of iterations, medians 32 and 178, means 38 and 38470.

Thus, with 99% confidence, these results indicate that **for area specifications for which a good diagram drawn with ellipses is known to exist**, euler*APE* **draws a good diagram**

- after the **first run** for 99.2% to 99.6% and after **1-10 reruns** for 99.9% to 100.0%;
- within **1 second** for 97.4% to 98.0% and within **10 seconds** for 99.6% to 99.8%.

Discussion and Examples

The results indicate that our **starting values the parameters of the optimization algorithm** ($p\gamma$, $pa\beta$ and $p\theta$, in Section 5.6.6) are appropriate for most drawable area specifications as good diagrams are generated for the majority during the first run. Also, rerunning the algorithm with starting values for its parameters that are 20% larger at every run was effective in obtaining a good diagram for those area specifications for which a non-good diagram was generated during the first run.

Figure 5.28 shows **examples of good diagrams** drawn after the first run for the area specifications obtained from the diagrams in DIAGLIB_*eval* shown in Figure 5.29. These examples illustrate that whenever possible eulerAPE draws circle-like curves (i.e., ellipses with similarly sized semi-axes). This is particularly apparent in Figure 5.28A, where the semi-axes of ellipses *a* and *b* are equal and those of *c* differ by only 2.8%, and Figure 5.28B, where the semi-axes of ellipses *a*, *b* and *c* differ by 6.0%, 5.4% and 8% respectively. However, in other cases as in Figure 5.28C and Figure 5.28D, elongated ellipses are required to accurately draw the required zone areas. Example, in Figure 5.28C, the required area for zones *c* and *abc* was small compared to that of zone *ac*, so ellipse *c* had to be elongated. In Figure 5.28D, the required area for the zones located in only one of the curves was large compared to that of other zones, so all the curves had to be elongated. We further discuss characteristics of area specifications that can be represented using ellipses and not circles in Section 5.7.2. Also, though no formal aesthetic criteria have been defined for such diagrams (Section 3.7.3), the diagrams generated by eulerAPE in Figure 5.28 could be considered more aesthetically desirable than those generated randomly in Figure 5.29. This is so, as most of the diagrams in Figure 5.28 have circle-like or similarly shaped curves (regular, convex curves with good continuity like circles are preferred, Section 3.7.3), that are often evenly distributed, easily identified and comparable (e.g., Figure 5.28A, D).
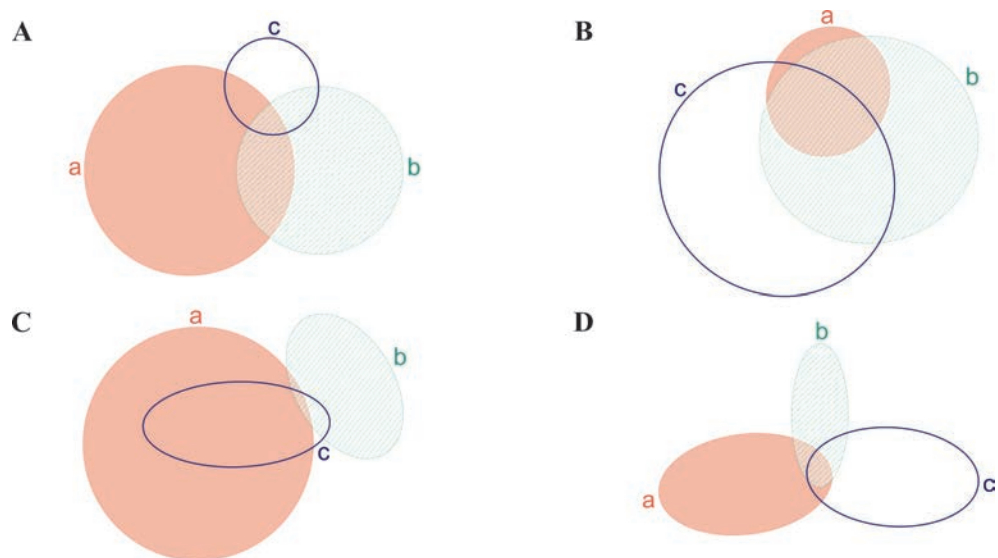
Figure 5.28: *Examples of good diagrams generated after the first run.* The area specifications, for which these diagrams were generated, were obtained from the diagrams in DIAGLIB_*eval* shown in Figure 5.29 and included: (A) {*a*=43555, *b*=23697, *c*=5870, *ab*=7393, *ac*=2129, *bc*=2051, *abc*=1106}, obtained from Figure 5.29A; (B) {*a*=2273, *b*=24458, *c*=44454, *ab*=7116, *ac*=740, *bc*=18807, *abc*=12092}, obtained from Figure 5.29B; (C) {*a*=53790, *b*=20052, *c*=177, *ab*=217, *ac*=20387, *bc*=1050, *abc*=444}, obtained from Figure 5.29C; (D) {*a*=17033, *b*=6248, *c*=16230, *ab*=615, *ac*=289, *bc*=840, *abc*=922}, obtained from Figure 5.29D. The time and number of iterations taken to generate these diagrams were: (A) 0.4 seconds, 44 iterations; (B) 0.3 seconds, 28 iterations; (C) 1.9 seconds, 123 iterations; (D) 0.7 seconds, 46 iterations.
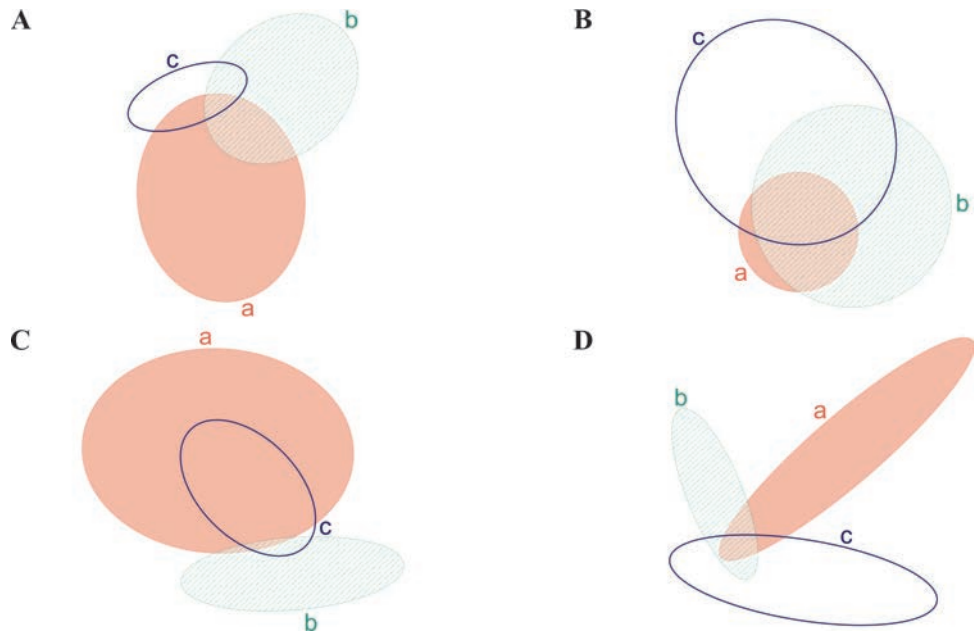
Figure 5.29: *Examples of diagrams in* DIAGLIB_*eval.* Good diagrams generated by euler*APE* for the area specifications of these diagrams are in Figure 5.28. The area specifications obtained from these diagrams are: (A) {*a*=43555, *b*=23697, *c*=5870, *ab*=7393, *ac*=2129, *bc*=2051, *abc*=1106}; (B) {*a*=2273, *b*=24458, *c*=44454, *ab*=7116, *ac*=740, *bc*=18807, *abc*=12092}; (C) {*a*=53790, *b*=20052, *c*=177, *ab*=217, *ac*=20387, *bc*=1050, *abc*=444}; (D) {*a*=17033, *b*=6248, *c*=16230, *ab*=615, *ac*=289, *bc*=840, *abc*=922}.
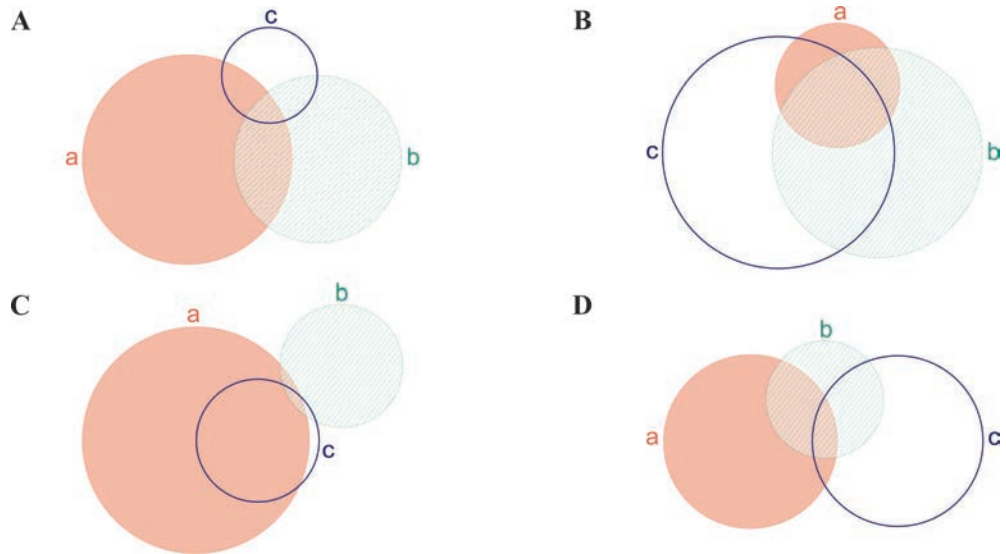
Figure 5.30: *The starting diagrams of the final good diagrams in Figure 5.28.* These were generated by euler*APE* for the area specifications obtained from the diagrams in DIAGLIB_*eval* shown in Figure 5.29, which were: (A) {*a*=43555, *b*=23697, *c*=5870, *ab*=7393, *ac*=2129, *bc*=2051, *abc*=1106}; (B) {*a*=2273, *b*=24458, *c*=44454, *ab*=7116, *ac*=740, *bc*=18807, *abc*=12092}; (C) {*a*=53790, *b*=20052, *c*=177, *ab*=217, *ac*=20387, *bc*=1050, *abc*=444}; (D) {*a*=17033, *b*=6248, *c*=16230, *ab*=615, *ac*=289, *bc*=840, *abc*=922}. The final good diagrams generated for the above area specifications using the starting diagrams in this figure were: (A) Figure 5.28A; (B) Figure 5.28B; (C) Figure 5.28C; (D) Figure 5.28D.

The generation of good diagrams in relatively fast time and with the above aesthetic features was possible through the generation of **rational starting diagrams** (Section 5.6.5). Figure 5.30 illustrates the starting diagrams used to generate the good diagrams in Figure 5.28. In some cases, the starting

diagram requires very few changes for it to be a good diagram, as Figure 5.30A with $diagError = 7.97 \times 10^{-4}$ which was used to generate the good diagram in Figure 5.28A. Generating starting diagrams that minimize the error of the zone in exactly the three ellipses, have helped the optimization avoid local minima, as good diagrams were eventually drawn for all the area specifications. Characteristics like the placement of the ellipses, where the largest two ellipses are placed at the bottom and vertically centre aligned, and the assignment of angles of rotation 0, $\pi/3$ and $2\pi/3$ for the three ellipses have contributed to the generation of well-balanced diagrams with evenly and harmoniously positioned curves, that are easily comparable (e.g., Figure 5.28A, D).

Generating **starting diagrams with ellipses whose semi-axis are equal to one another such that the ellipses are circles**, increases the likelihood that the curves in the generated good diagram are circle-like (e.g., Figure 5.28A, B). In fact, Figure 5.31 demonstrates that the majority of the generated diagrams had ellipses with ratios of their semi-axes close to one. Out of the 10,000 diagrams, 59.4% and 34.7% had at least one ellipse with semi-axes differing by respectively 10% and 5% (i.e., a ratio of semi-axes in [0.90,1.10] for 10% and [0.95,1.05] for 5%), 20.6% and 6.5% of the diagrams had at least two ellipses with semi-axes differing by respectively 10% and 5%, and 3.4% and 0.4% of the diagrams had all the three ellipses with semi-axes differing by respectively 10% and 5%. As illustrated in Figure 5.28B, it is not easily noted that ellipses with semi-axes differing by 10% or less are not circles. Circles are preferred (Section 3.7.3), but they cannot depict most zone areas accurately (Section 3.7.6). So using such ellipses when possible, accurate diagrams with circle-like curves can be generated. These results show the effectiveness of our starting diagrams in helping the optimization avoid local minima and generate good diagrams in relatively fast time, with aesthetic features that are often desirable.

Figure 5.31: *Distributions and boxplots of the ratio of the semi-axes of ellipses a, b and c of the 10,000 good diagrams generated for the area specifications obtained from* DIAGLIB_eval. *The distributions illustrate the ratios of the semi-axes of the ellipses in only [0, 4], as very few diagrams had greater ratios (medians 1.01, 0.99, 1.00 for respectively ellipses a, b, c; means 1.07, 1.04, 1.06 for respectively ellipses a, b, c). Black bars indicate those diagrams whose ratio of the semi-axes of the three ellipses was in [0.90, 1.10]. If the ratio of the semi-axes is one, then the ellipse is a circle.*

As mentioned earlier, the majority (88.5%) of **the non-good diagrams generated in the first run** had a low *diagError* and only needed further refinement. Figure 5.32A is an example of such a diagram generated during the first run with $diagError = 6.51 \times 10^{-4}$. Figure 5.32B is the good diagram that was generated after one rerun, when the starting values of the parameters of the optimization algorithm were

set to $p\gamma$ = 12 pixels, $p\alpha\beta$ = 6% and $p\theta$ = 4π/5. This good diagram was generated in a total of 1.2 seconds and 86 iterations (including the first run and one rerun) for the area specifications obtained from the diagram in DIAGLIB_*eval* in Figure 5.32C. Figure 5.32B might be considered more aesthetically desirable than Figure 5.32C, as the curves are circle-like (ellipses *b* and *c* have equal semi-axes and are thus circles, while ellipse *a* has semi-axes differing by 3.3%) and positioned evenly and harmoniously, in a way that all the curves are easily comparable. The area of zone *bc* is 0.003% of the total diagram area. However, with our analytic method for computing the area of overlapping ellipses (Section 5.4), eulerAPE is still capable of accurately computing such zone areas and generating a good diagram. This meets our objective and Tufte's graphical integrity principle [Tufte, 1983] in generating diagrams with zone areas that are accurate and directly proportional to the required quantitative data (Section 5.2.3).
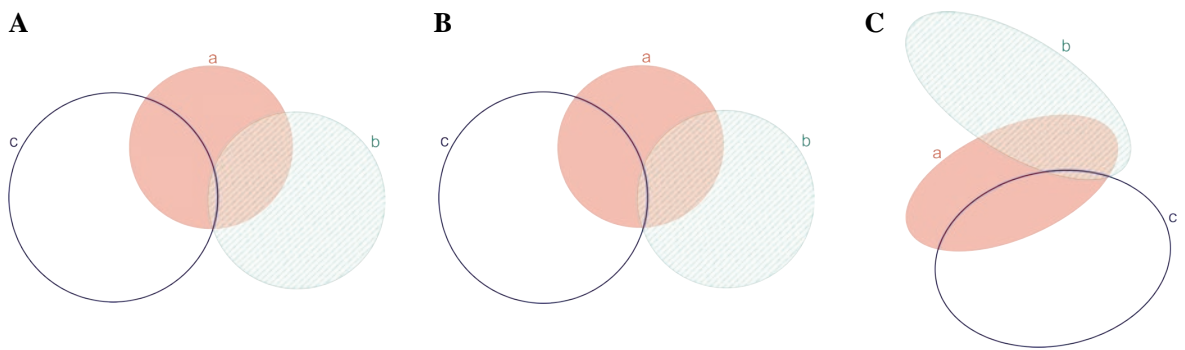
Figure 5.32: *An example of a non-good diagram with a very low diagError drawn after run 1 and the good diagram drawn after rerun 1, for the area specification obtained from a diagram in* DIAGLIB_*eval*. (A) The non-good diagram drawn after run 1 (*diagError* = 6.51×10⁻⁴) and (B) the good diagram drawn after rerun 1, for {*a*=10018, *b*=27132, *c*=39737, *ab*=9567, *ac*=11454, *bc*=3, *abc*=668} obtained from (C) a diagram in DIAGLIB_*eval*. B was drawn in 1.2 seconds and 86 iterations (including run 1 and the one rerun).

It could be argued that such **small zones are not visible** and could be considered missing. However, no empirical evidence indicates how the zone areas are perceived and what areas are not perceived (Section 3.7.3). Also, area specifications for which a good diagram can be drawn with ellipses should be identified, so zone area accuracy is important. If users of eulerAPE do not need to depict such small areas and a **3-Euler diagram** is required, the area of zones of the 3-Venn diagram that are not required could be set to a value close to zero, such that the zones are still drawn accurately, but they are too small to be visible on the screen. Out of the 10,000 area specifications for which a good diagram was generated, 8 (0.08%), 543 (5.4%), 2355 (23.6%) and 6898 (69.0%) had at least one zone with an area that was respectively 0.001%, 0.01%, 0.1% and 1% of the total required diagram area. Good diagrams were drawn in the first run for the majority of these area specifications (75.0%, 94.1%, 97.7%, 99.1% of those with at least one zone area that was respectively 0.001%, 0.01%, 0.1%, 1% of the total required diagram area). Being able to handle zones with an area close to zero suggests that our algorithm is also effective in drawing Euler diagrams (not only Venn diagrams), but further evaluation has to be carried out.

Figure 5.33B is **another example of a good diagram drawn for an area specification with a few very small zone areas** compared to others in the required diagram. The area specification was obtained from the diagram in DIAGLIB_*eval* in Figure 5.33C. The required area for both zone *c* and zone *bc* was 0.003% of the total diagram area. Thus the reason why Figure 5.33B seems to be an Euler diagram even though it is a wellformed Venn diagram. This good diagram was generated during the second rerun, within a total of 49.5 seconds and 4785 iterations (including the first run and the two reruns). In the

previous runs, a local minimum was reached where, as shown in Figure 5.33A (the non-good diagram generated during the first run) ellipse *c* had a circle-like shape and ellipse *a* was tilted to the left leading to a point where any change to the diagram would make it non-wellformed or not a Venn diagram. Rerunning the optimization algorithm with its parameters assigned the starting values $p\gamma = 14.4$ pixels, $pa\beta = 7.2\%$ and $p\theta = 24\pi/25$, the exploration of the search space was widened and the good diagram was found. At first glance, zone *a* in Figure 5.33C seems to be made up of two minimal regions and so, Figure 5.33B could be preferred, but no empirical evidence supports this observation (Section 3.7.3).

Figure 5.34B is **an example of a good diagram that was generated after the first rerun** for the area specification obtained the diagram in DIAGLIB_*eval* in Figure 5.34C. As shown in Figure 5.34A (the non-good diagram generated during the first run), during the first run, the optimization was trapped in a local minimum as ellipse *b* approached the edge of ellipse *c* (making zone *c* seem like it was made up of two minimal regions) and no change that would keep the diagram wellformed and a Venn diagram was possible. Rerunning the optimization with parameters $p\gamma = 12$ pixels, $pa\beta = 6\%$, $p\theta = 4\pi/5$ at the start, Figure 5.34B was generated. Since in Figure 5.34C, the space between the edge of ellipses *b* and *c* in zone *b* is narrow and not easily visible, Figure 5.34B could be preferred. However, the effect of the shape of the curves and zoneson diagram comprehension have not been studied empirically and thus, in contrast to zone area accuracy, these aesthetic preferences are currently subjective (Section 3.7.3).
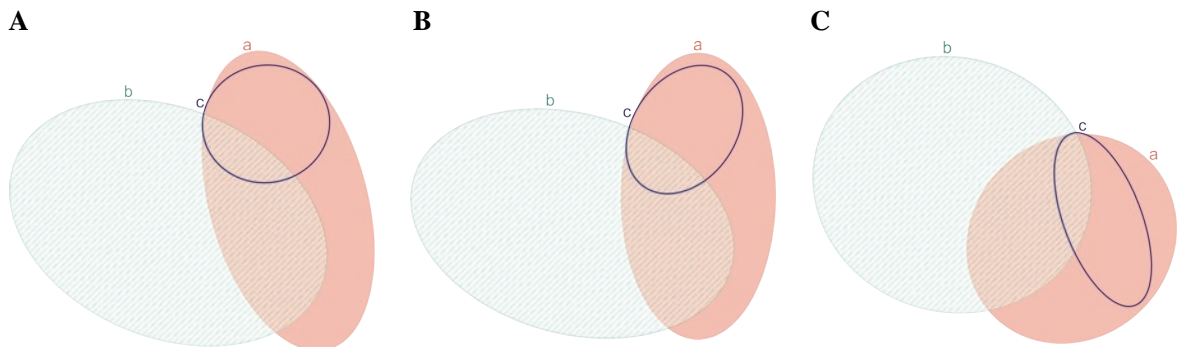
Figure 5.33: *An example of a non-good diagram generated during the first run and the good diagram generated during the second rerun, for the area specification obtained from a diagram in* DIAGLIB_*eval.* (A) The non-good diagram drawn in run 1 (*diagError* = 4.03×10^{-1}) and (B) the good diagram drawn after rerun 2, for {*a*=17671, *b*=65298, *c*=3, *ab*=19904, *ac*=12951, *bc*=4, *abc*=4816} obtained from (C) a diagram in DIAGLIB_*eval*. B was generated in 49.5 seconds and 4785 iterations (including run 1 and the two reruns).
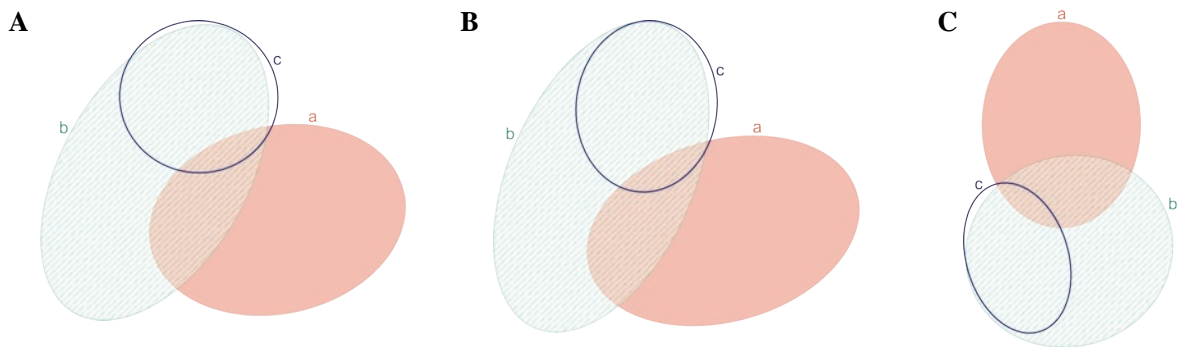
Figure 5.34: *An example of a non-good diagram generated during the first run and the good diagram generated during the first rerun, for the area specification obtained from a diagram in* DIAGLIB_*eval.* (A) The non-good diagram drawn after run 1 (*diagError* = 8.38×10^{-3}) and (B) the good diagram drawn after rerun 1, for {*a*=53804, *b*=39550, *c*=1256, *ab*=15606, *ac*=15, *bc*=29904, *abc*=3597} obtained from (C) a diagram in DIAGLIB_*eval*. B was generated in 2.9 seconds and 367 iterations (including run 1 and the one rerun).

This evaluation demonstrates that eulerAPE always draws a good diagram for an area specification that is drawable. So if eulerAPE cannot draw a good diagram, then an accurate wellformed 3-Venn diagram drawn with ellipses does not exist for that area specification. Thus using eulerAPE, we can assess the type and percentage of all area specifications for which a good diagram can be drawn with ellipses.

5.7.2 For Any Area Specification

To evaluate the effectiveness of ellipses in drawing accurate area-proportional 3-Venn diagrams for any area specification, we used eulerAPE to draw diagrams using ellipses for the 10,000 area specifications in the AREASPECLIB_*eval*. Each area specification was made of seven random values, indicating the zone areas for a 3-Venn diagram. So in contrast to the area specifications in Section 5.7.1, it was unknown if these area specifications were drawable and could be depicted accurately with a diagram drawn with ellipses. The library and the characteristics of the area specifications are discussed in Section 5.5.

Since circles are often preferred for these diagrams (Section 3.7.3), we verified whether, for any of the area specifications in AREASPECLIB_*eval*, a good diagram could be drawn using circles rather than ellipses. To do so, we generated diagrams for these area specifications using a variant of eulerAPE that restricts all the ellipses in the diagrams to circles, such that each individual ellipse has equal semi-axes. These results were then used to identify characteristics of area specifications for which a good diagram can or cannot be drawn using ellipses and others for which circle-like ellipses or circles could be used.

When using eulerAPE, the rerun option of the optimization algorithm was enabled to ensure that a good diagram is drawn for all drawable area specifications (Section 5.7.1). A maximum number of 10 reruns are allowed (Section 5.6.6). If by the tenth rerun a good diagram is not found, eulerAPE returns the diagram with the lowest *diagError* out of the 11 generated in total during the first run and the 10 reruns.

Results and Discussion

Good diagrams drawn with ellipses were generated for 8607 of the 10,000 area specifications (i.e., 86.1%)—8372 after the first run (i.e., 97.3% of the 8607) and 235 after one to a maximum of 10 reruns (i.e., 2.7% of 8607). Figure 5.35 illustrates that more than half of the 235 good diagrams (56.2%) were generated during the first rerun and only 1 was generated after 10 reruns, as the *diagError* of the non-good diagrams generated for these area specifications during the first run was relatively low (*diagError* in [1.51×10^{-6}, 3.28×10^{-2}] with median 1.89×10^{-3} and mean 3.77×10^{-3}).
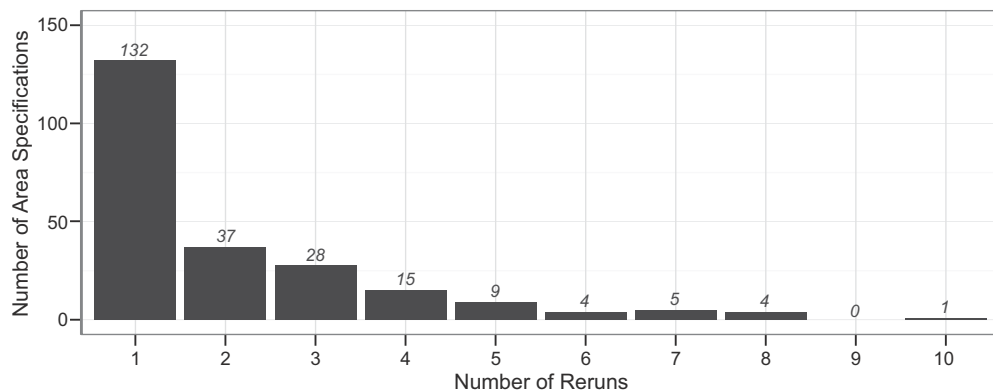
Figure 5.35: *The number of reruns (1-10) for* eulerAPE *to draw a good diagram for the 235 area specifications obtained from* AREASPECLIB_*eval for which a non-good diagram was drawn after run 1.*

None of the diagrams drawn with circles for the 10,000 area specifications were good. The *diagError* of these diagrams was greater than that of the non-good diagrams drawn with ellipses (Figure 5.36). In all of these cases, the optimization was rerun 10 times and a non-good diagram was generated at every run. The diagram with the lowest *diagError* was then returned. So the 1393 non-good diagrams with ellipses and the 10,000 with circles are the most accurate diagrams that were generated for the required area specification. Most of these diagrams drawn with ellipses were generated during the initial reruns (Figure 5.37). Similar percentages of the diagrams drawn with circles were generated during each of the 10 reruns. So with ellipses, fewer reruns than with circles are required to get a diagram that is closest to the required out of all the ones found by euler*APE* during the various runs.
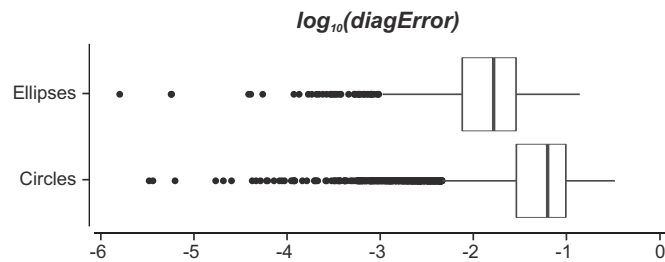
Figure 5.36: *The $\log_{10}(diagError)$ of the non-good diagrams generated for the area specifications in* AREASPECLIB_*eval using ellipses and circles.* Out of the 10,000 area specifications, non-good diagrams were generated for 1393 using ellipses and 10,000 using circles. *diagError*: ellipses in $[1.59\times10^{-6}, 1.39\times10^{-1}]$, median 1.65×10^{-2}, mean 2.11×10^{-2}; circles in $[3.30\times10^{-6}, 3.31\times10^{-1}]$, median 6.28×10^{-2}, mean 6.73×10^{-2}. A non-good diagram is the diagram with the lowest *diagError* that is generated during the 11 runs.
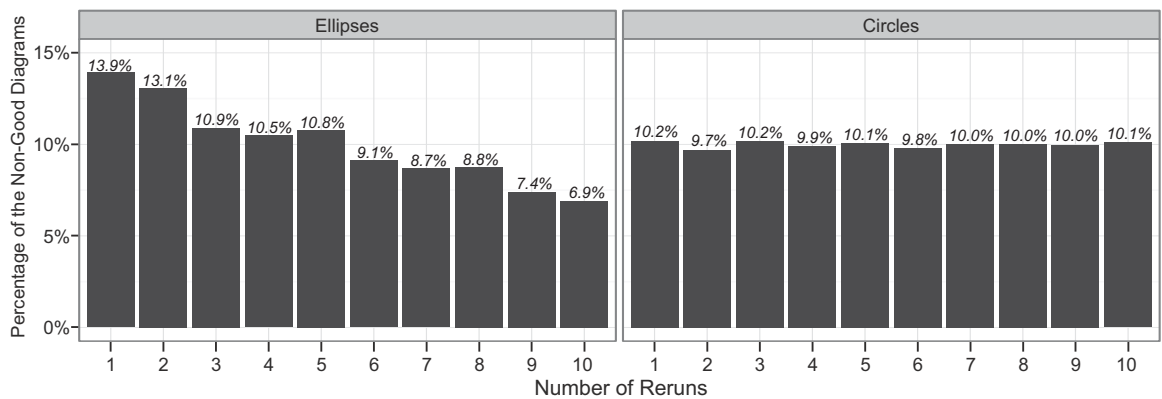
Figure 5.37: *The reruns (1-10) during which the non-good diagrams returned by* euler*APE and drawn using ellipses and circles were generated—percentages of the total number of 1393 non-good diagrams drawn with ellipses and 10,000 non-good diagrams drawn with circles are shown.*

Thus, with 99% confidence, these results indicate that **for any area specification corresponding to the abstract description of a 3-Venn diagram**, **a good diagram** can be drawn (using euler*APE*)

- with **ellipses** for 85.2% to 86.9% and with **circles** for 0.0% to 0.1%;

There are area specifications for which an accurate area-proportional wellformed 3-Venn diagram cannot be drawn using convex curves (Section 3.7.5), so drawing good diagrams with ellipses for a large majority of the 10,000 random area specifications in AREASPECLIB_*eval*, indicates great potential for using curves that are regular and smooth like circles, but with more degrees of freedom like ellipses.
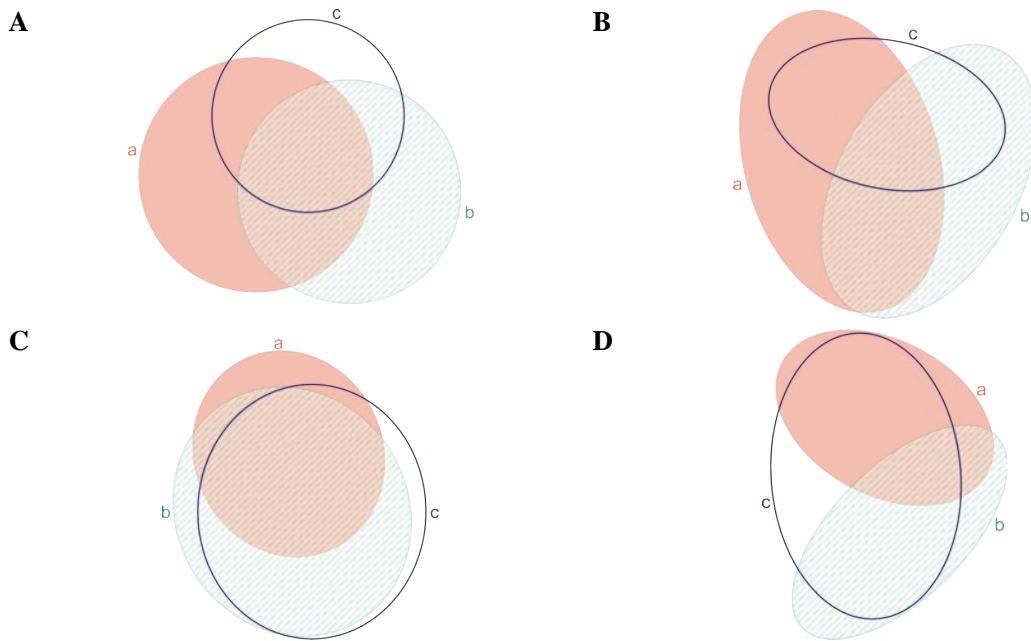
Figure 5.38: *Examples of good and non-good diagrams generated for area specifications in* AREASPECLIB*_eval using either circles or ellipses.* (A) A non-good diagram generated using circles for {*a*=8698, *b*=7937, *c*=424, *ab*=7488, *ac*=6484, *bc*=4398, *abc*=4768}, with *diagError*=0.08 and zone areas {*a*=10985, *b*=10153, *c*=4164, *ab*=4988, *ac*=4101, *bc*=2460, *abc*=8361}. (B) A good diagram generated using ellipses for {*a*=8698, *b*=7937, *c*=424, *ab*=7488, *ac*=6484, *bc*=4398, *abc*=4768}, same area specification as in A. The ratio of the semi-axes (semi-minor axis / semi-major axis) for ellipses *a*, *b* and *c* are respectively 0.63, 0.60, 0.62. Thus, the semi-axes of the ellipses differ by ~40%. (C) A good diagram generated using ellipses for {*a*=1783, *b*=1311, *c*=1370, *ab*=771, *ac*=414, *bc*=6820, *abc*=8915}. The ratio of the semi-axes for ellipses *a*, *b* and *c* are respectively 90.0, 0.91, 0.90. Thus, the semi-axes of the ellipses differ by 10%. (D) A non-good diagram generated using ellipses for {*a*=618, *b*=1612, *c*=891, *ab*=978, *ac*=6979, *bc*=6255, *abc*=887}, with *diagError* = 0.07 and zone areas {*a*=1194, *b*=1967, *c*=1824, *ab*=663, *ac*=5612, *bc*=4949, *abc*=1907}.

Figure 5.38 illustrates **examples of good and non-good diagrams** generated with circles or ellipses. For the same area specification, a non-good diagram Figure 5.38A was generated with circles and a good diagram Figure 5.38B was generated with ellipses. The diagram in Figure 5.38A is misleading. Despite having the smallest area in the area specification, zone *c* is bigger than other zones like zone *bc* and *ac*. Zone *abc* should be similar in size to zone *bc* and smaller than zones *ab* and *ac*, yet is bigger than the three of them. These problems are not apparent in Figure 5.38B as the zone areas are accurate, the curves are easily identified and the ellipses have a similar shape (all the ellipses have semi-axes differing by 40%). Figure 5.38C is also a good diagram with ellipses, but here the ellipses have a more circle-like and similar shape as their semi-axes differ by 10%. Figure 5.38D is an non-good diagram with ellipses, as the required areas for the zones in only two curves are much larger than those in only one curve. Such area specifications are in fact non-drawable with ellipses (discussed in next subsection).

For drawable area specifications, eulerAPE (and ellipses) often generates **diagrams with circle-like ellipses** (Section 5.7.1). For random area specifications for which a good diagram drawn with ellipses might not even exist, fewer diagrams had circle-like ellipses with their semi-axes differing by at most 10% (as Figure 5.38C), and more diagrams had ellipses with semi-axes differing by 20% to 40% (as Figure 5.38B). This is apparent in Figure 5.39, which shows the distributions and boxplots of the ratio of the semi-axes of the ellipses in the 8607 good and 1393 non-good diagrams generated by eulerAPE.

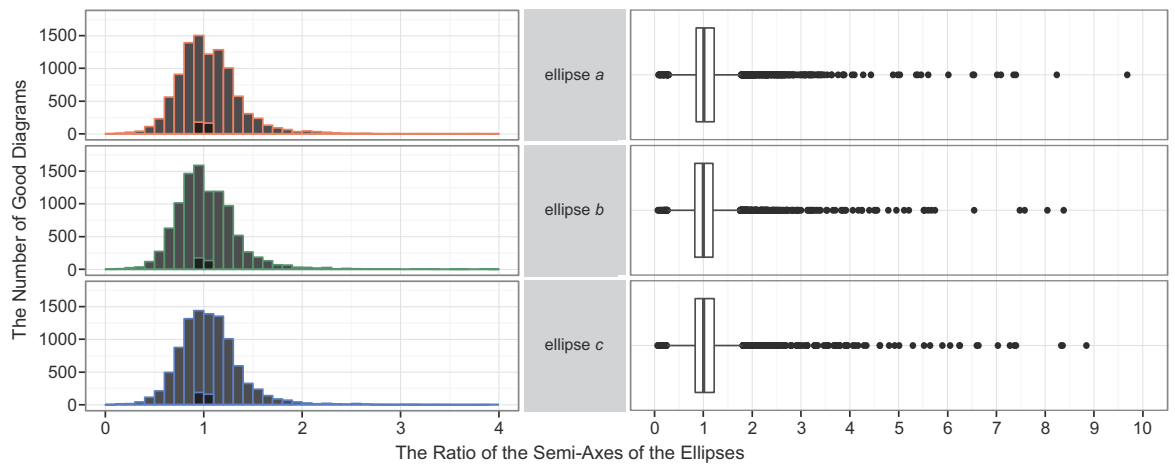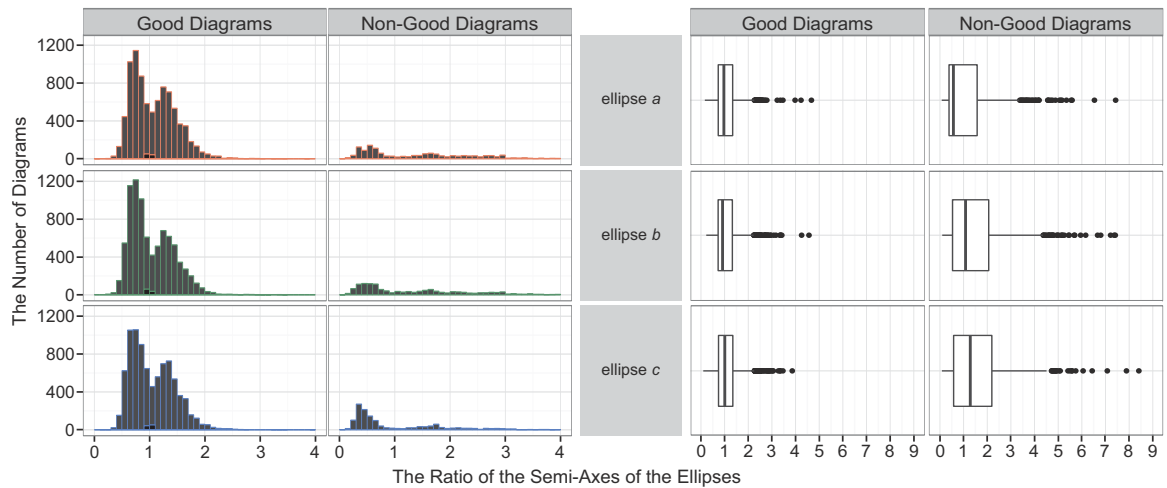**Figure 5.39**: *Distributions and boxplots of the ratio of the semi-axes of ellipses a, b and c of the 8607 good diagrams and the 1393 non-good diagrams generated with ellipses for the area specifications in* AREASPECLIB_*eval.* The distributions illustrate the ratios of the semi-axes of the ellipses in only [0,4], as very few diagrams had greater ratios (for ellipses a, b, c: good diagrams, medians 0.97, 0.91, 1.01, means 1.05, 1.04, 1.07; non-good diagrams, medians 0.57, 1.09, 1.28, means 1.02, 1.43, 1.49 for ellipses a, b, c). Black bars indicate those diagrams whose ratio of the semi-axes of the three ellipses was in [0.90, 1.10]. If the ratio of the semi-axes is one, then the ellipse is a circle.

The time and number of iterations that were required for the generation of the good diagrams using ellipses were similar to those of our evaluation in Section 5.7.1 for drawable area specifications (median 0.4 seconds, 35 iterations; mean 1.9 seconds, 201 iterations). As expected, non-good diagrams required more time and iterations as the optimization was rerun 10 times (median 4.0 seconds, 586 iterations; mean 25.9 seconds, 4417 iterations). All the diagrams drawn with circles required more time and iterations as they were all non-good. Yet, their medians (3.2 seconds, 500 iterations) and means (3.4 seconds, 529 iterations) were lower than those of non-good diagrams with ellipses, possibly because the diagram search space for circles is smaller than that for ellipses, and fewer of the explored changes during the search reduce its cost since circles cannot draw most diagrams accurately (Section 3.7.6).

Similar to the results obtained in our evaluation in Section 5.7.1, the majority of the 10,000 diagrams with ellipses were generated within 1 second (84.1%: 8405/8607 good, 0/1393 non-good) and nearly all of the 10,000 diagrams with ellipses within 10 seconds (96.9%: 8569/8607 good, 1119/1393 non-good). These results also indicate that eulerAPE does not return a diagram within 1 second for any of the area specifications that are non-drawable, but it does return a diagram within 10 seconds for 77.4% to 82.9% of the non-drawable area specifications (99% confidence). Also, out of the 10,000 diagrams with circles, none were generated eulerAPE 1 second, but 99.6% (9959/10,000) were generated eulerAPE 10 seconds

Thus, with 99% confidence, these results indicate that **for any area specification corresponding to the abstract description of a 3-Venn diagram**, a diagram can be drawn with ellipses (using eulerAPE)

- within **1 second** for 83.1% to 85.0% and within **10 seconds** for 96.4% to 97.3%, irrespective to whether the diagram is **good or non-good**;

- within **1 second** for 97.2% to 98.0% and within **10 seconds** for 99.3% to 99.7%, when the diagram is **good**.

Characteristics of Drawable and Non-Drawable Area Specifications

eulerAPE always draws a good diagram for drawable area specifications (Section 5.7.1). This means that the 8607 area specifications in AREASPECLIB_*eval* for which eulerAPE drew a good diagram with ellipses are drawable, while the remaining 1393 area specifications for which eulerAPE drew a non-good diagram with ellipses are non-drawable. By analysing these area specifications (as in Figure 5.40), we identified characteristics of area specifications for which a good diagram can or cannot be drawn using ellipses.
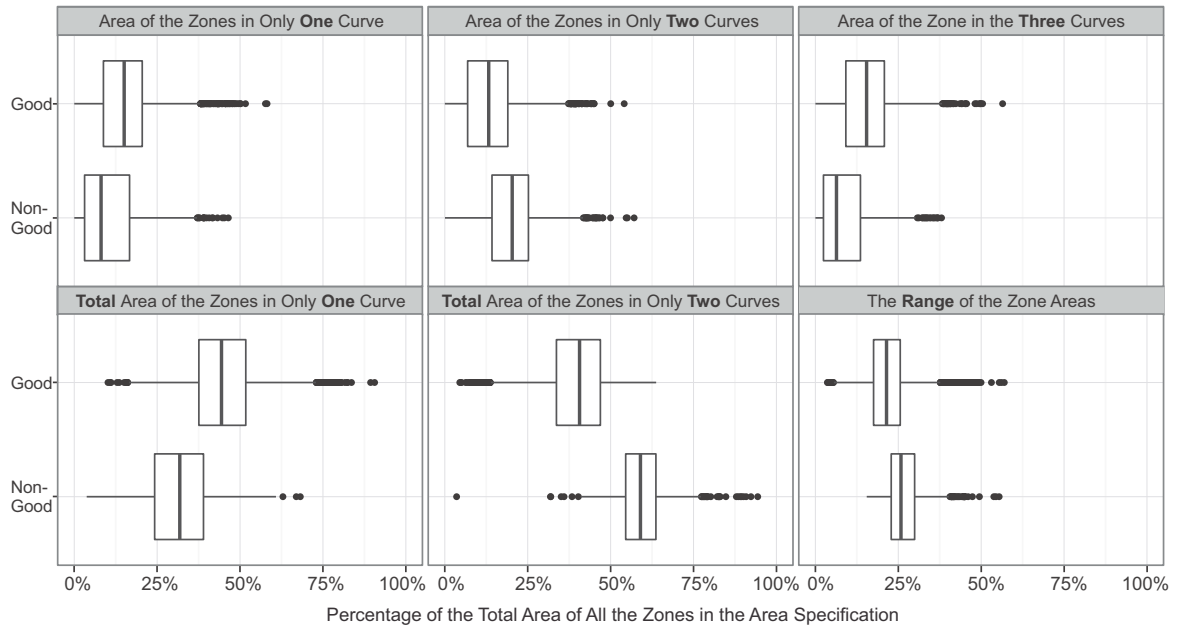
Figure 5.40: *The characteristics of the area specifications in* AREASPECLIB_*eval for which a good or a non-good diagram was generated using ellipses (and* eulerAPE*).* Out of the 10,000 generated diagrams, 8607 were good and 1393 were non-good with respect to the area specifications in AREASPECLIB_*eval*. The areas are shown as percentages of the total area of all the zones in their respective area specification.

Table 5.1: *The median and mean area percentages with respect to the diagram area for the 1-curve zones, 2-curve zones and 3-curve zone of the 10,000 area specifications in* AREASPECLIB_*eval that are drawable (8607) and non-drawable (1393) with ellipses (and* eulerAPE*).*

	drawable			non-drawable		
	1-curve zones	2-curve zones	3-curve zone	1-curve zones	2-curve zones	3-curve zone
median	15.1%	13.2%	15.5%	8.1%	20.3%	6.4%
mean	14.9%	13.3%	15.3%	10.6%	19.8%	8.9%

   The boxplots in Figure 5.40 illustrate the required area for different types and groups of zones as percentages of the diagram area, of the 10,000 area specifications in AREASPECLIB_*eval* for which diagrams were drawn with ellipses (and eulerAPE). In this section, we will refer to zones in only *n* curves as *n*-curve zones. These boxplots and Table 5.1 indicate that **drawable area specifications often had: (1) larger areas for the 1-curve zones than for the 2-curve zones; (2) an area for the 3-curve zone that is the largest of all the zone areas, but similar to the areas for the 1-curve zones**. Also, 58.1% of the 8607 drawable area specifications (5004/8607) had a greater total area for the 1-curve zones (median 44.4%, mean 44.8%, of diagram area) than for the 2-curve zones (median 40.6%, mean 39.9%, of diagram area). The other 41.9% of the 8607 drawable area specifications (3603/8607) had a total area

for the 1-curve zones that was smaller than that for the 2-curve zones by median and mean factors of 10.5% and 12.3%, and an area for the 3-curve zone (median 16.1%, mean 16.1%, min 0.3%, of diagram area) that was larger than that of the other 5004 drawable area specifications (median 14.8%, mean 14.7%, min 0.01%, of diagram area). These 3603 drawable area specifications also indicated that the greater (a) the difference between the total area for the 1-curve zones and the total area for the 2-curve zones, the greater (b) the area for the 3-curve zone (median and mean of $b\,/a$ were 1.38 and 7.06).

Figure 5.41 shows two **examples of good diagrams** generated for 2 of the 8607 area specifications: (A) with the total area for the 1-curve zones larger than the total area for the 2-curve zones, and with zone area percentages (with respect to the diagram area) that differed by 5% from the median and mean for drawable area specifications (Figure 5.41A); (B) with the total area for the 1-curve zones smaller than the total area for the 2-curve zones, and with a large area for the 3-curve zone (Figure 5.41B; Figure 5.38C is also of type B, whose area for the 3-curve zone is the largest zone).
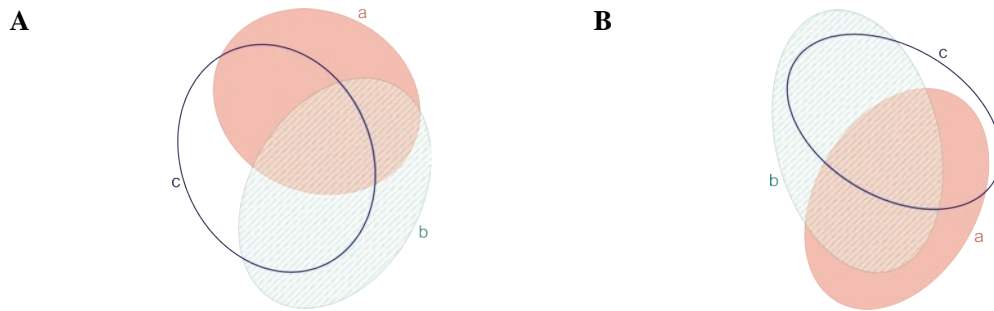
Figure 5.41: *Examples of good diagrams generated for area specifications in* AREASPECLIB_*eval.* (A) A good diagram generated for {a=6330, b=8151, c=5911, ab=3730, ac=6851, bc=7670, abc=7503}. The zone areas as percentages of the total area of all the zones in this area specification are {a=13.7%, b=17.7%, c=12.8%, ab=8.1%, ac=14.8%, bc=16.6%, abc=16.3%}, where the total area of zones a, b and c is 44.2% of the diagram area and the total area of zones ab, ac and bc is 39.6% of the diagram area with range of the zone area percentages of 9.6%. This is an example of a drawable area specification with the total area of the 1-curve zones larger than the total area of the 2-curve zones, and with zone area percentages with respect to the diagram area that differed by 5% from the median and mean for drawable area specifications. (B) The good diagram generated for {a=6266, b=5170, c=2232, ab=7256, ac=4218, bc=8684, abc=8458}. The zone areas as percentages of the total area of all the zones in this area specification are {a=14.8%, b=12.2%, c=5.3%, ab=17.2%, ac=10.0%, bc=20.5%, abc=20.0%}, where the total area of zones a, b and c is 32.3% of the diagram area and the total area of zones ab, ac and bc is 47.7% of the diagram area with range of the zone area percentages of 15.3%. This is an example of a drawable area specification with the total area of the 1-curve zones smaller than the total area of the 2-curve zones, and with a large area for the 3-curve zone.

The boxplots (Figure 5.40) and Table 5.1 indicate that **the opposite was true for most of the non-drawable area specifications** for which a non-good diagram was generated, as the areas for the 1-curve zones were often smaller than those for the 2-curve zones, and the 3-curve zone often had the smallest area of all the zones, which area was similar in quantity to the areas for the 1-curve zones. Also, 95.8% of the 1393 non-drawable area specifications (1335/1393) had a smaller total area for the 1-curve zones (median 31.8%, mean 31.7%, of diagram area) than for the 2-curve zones (median 59.0%, mean 59.5%, of diagram area). In such cases, the greater (a) the difference between the total area for the 1-curve zones and the total area for the 2-curve zones, the smaller (b) the area for the 3-curve zone (median and mean of b/a were 0.26 and 0.46). The other 4.2% of the 1393 non-drawable area specifications (58/1393) had a total area for the 1-curve zones that was larger than that for 2-curve zones by median and mean

factors of 6.1% and 8.1%, and an area for the 3-curve zone (median 1.0%, mean 2.6%, max 28.3%, of diagram area) that was smaller than that of the other 1335 non-drawable area specifications (median 6.7%, mean 9.1%, max 38.1%, of diagram area). In fact, all except for six of these 58 non-drawable area specifications had an area for the 3-curve zone that was less than 6% of diagram area.

Figure 5.42 shows **examples of non-good diagrams** generated for 2 of the 1393 area specifications: (A) with the total area for 1-curve zones smaller than the total area for 2-curve zones, with zone area percentages (with respect to diagram area) that differed by 5% from median and mean for non-drawable area specifications (Figure 5.42A); (B) with the total area for 1-curve zones larger than that for 2-curve zones, but with a very small area for the 3-curve zone and for one of the 1-curve zones (Figure 5.42B).

A

B
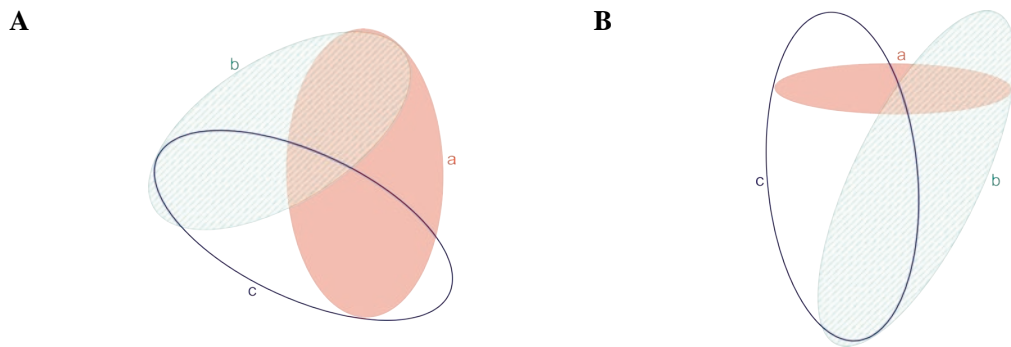
Figure 5.42: *Examples of non-good diagrams generated for area specifications in* AREASPECLIB_*eval.* (A) The non-good diagram generated for {*a*=3300, *b*=4000, *c*=4043, *ab*=5808, *ac*=7598, *bc*=5800, *abc*=995} with *diagError* = 0.02. The zone areas as percentages of the total area of all the zones in this area specification are {*a*=10.5%, *b*=12.7%, *c*=12.8%, *ab*=18.4%, *ac*=24.1%, *bc*=18.4%, *abc*=3.2%}, where the total area of zones *a*, *b* and *c* is 36.0% of the diagram area and the total area of zones *ab*, *ac* and *bc* is 60.9% of the diagram area with range of the zone area percentages of 20.9%. This is an example of a non-drawable area specification with the total area of the zones in only one curve is smaller than the total area of the zones in only two curves and with zone area percentages with respect to the diagram area that were 5% smaller or greater than the median and mean for non-drawable area specifications. (B) The non-good diagram generated for {*a*=31, *b*=7944, *c*=9959, *ab*=2297, *ac*=2834, *bc*=8114, *abc*=1} with *diagError* = 0.01. The zone areas as percentages of the total area of all the zones in this area specification are {*a*=0.1%, *b*=25.5%, *c*=31.9%, *ab*=7.4%, *ac*=9.1%, *bc*=26.0%, *abc*=0.003%}, where the total area of zones *a*, *b* and *c* is 57.5% of the diagram area and the total area of zones *ab*, *ac* and *bc* is 42.5% of the diagram area with range of the zone area percentages of 31.9%. This is an example of a non-drawable area specification with the total area of the zones in only one curve greater than the total area of the zones in only two curves and with a very small area for the zone in exactly the three curves and for one of the zones in only one curve.

The boxplots (Figure 5.40) also show that **the ranges of the zone area percentages** of drawable area specifications were smaller (median 21.5%, mean 21.8%) than those of non-drawable area specifications (median 25.8%, mean 26.9%), indicating that the greater the difference between the largest and smallest zone area, the greater the likelihood that the area specification is non-drawable.

Earlier we noted that **the 3-curve zone is typically one of the largest zones in drawable area specification and one of the smallest zones in non-drawable area specifications**. This is evident in Figure 5.43 which shows the area of the 3-curve zone in the 10,000 area specifications as a ratio of: the mean area of 1-curve zones, the mean area of 2-curve zones, the largest area of all the zones. These boxplots indicate that drawable area specifications often had an area for the 3-curve zone that was: greater than the mean area of the 1-curve zones (ratio in [0.0006,13.5] with median 1.05, mean 1.17)

and 2-curve zones (ratio in [0.001, 21.4] with median 1.13, mean 1.32); similar in quantity to the largest zone area in the area specification (ratio: median 0.65, mean 0.62). For 16.2% of the 8607 area specifications (i.e., 1393/8607), the 3-curve zone was the largest zone in the area specification (as in Figure 5.38C) and for 11.9% (i.e., 1024/8607), it was the smallest zone in the area specification. Non-drawable area specifications often had an area for the 3-curve zone that was: greater than the mean area of the 1-curve zones (ratio in [0.0002, 27.8] with median 0.59, mean 1.24), but smaller than the mean area of the 2-curve zones (ratio in [0.0002, 23.8] with median 0.32, mean 0.48); dissimilar in quantity to the largest zone area in the area specification (ratio: median 0.22, mean 0.62). For 3.1% of the 1393 area specifications (i.e., 43/1393), the 3-curve zone was the largest zone in the area specification, and for 27.9% (i.e., 389/1393), it was the smallest zone in the area specification. For both the drawable and non-drawable area specification, the area of the 3-curve zone was greater than the smallest zone area in the area specification by a mean factor of 19 and 28 respectively.
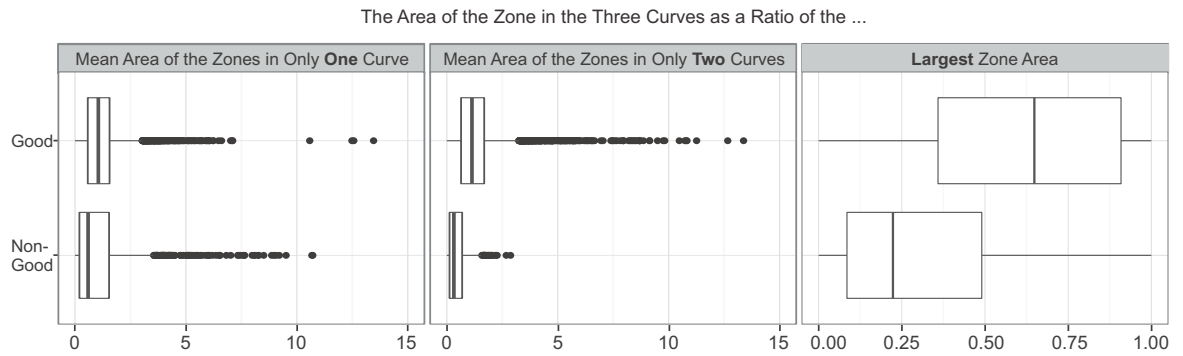
Figure 5.43: *The area of the zone in exactly the three curves as a ratio of the mean area of the zones in only one curve, the mean area of the zones in only two curves, and the largest zone area, for all the area specifications in* AREASPECLIB_*eval for which a good or a non-good diagram was generated using ellipses (and* eulerAPE*).* Out of the 10,000 generated diagrams, 8607 were good and 1393 were non-good with respect to the area specifications in AREASPECLIB_*eval*.

Thus, **a drawable area specification often has: (1) larger areas for the 1-curve zones than for the 2-curve zones, and (2) an area for the 3-curve zone that is similar to those for the 1-curve zones**. (1) is consistent with the characteristics of the area specifications obtained from DIAGLIB type libraries (DIAGLIB_*design*, DIAGLIB_*eval*) for which a good diagram is known to exist (Section 5.5.2).

We proceed with a comparative evaluation of the effectiveness of ellipses, circles and polygons, and various drawing methods including eulerAPE, in drawing accurate, easy to comprehend area-proportional 3-Venn diagrams for randomly generated area specifications and for real world data. We start by comparing diagrams drawn using ellipses and eulerAPE with those drawn using circles and venneuler.

5.7.3 Comparison with Circles and venneuler

Using a variant of eulerAPE, our evaluation in Section 5.7.2 indicates that it is highly unlikely that there area specifications corresponding to a 3-Venn diagram for which a good diagram can be drawn with circles. To verify this finding, we used the latest circle-based method venneuler [Wilkinson, 2012] version 1.1-0 to generate diagrams with circles for the 10,000 area specifications in AREASPECLIB_*eval*. We then compared the accuracy of venneuler's diagrams with the accuracy of the diagrams generated by eulerAPE with circle and ellipses in Section 5.7.2.

The method `venneuler` aims at generating zone areas that are accurate and directly proportional to the required quantitative data, without considering any diagram aesthetics not even wellformedness. It is currently considered the most effective in drawing accurate diagrams with circles. It is the only method to take a statistical approach and differs from `eulerAPE` in various ways as discussed in Section 5.6, Section 3.7.4. Example, `venneuler` uses a numerical approximation method to compute the zone areas and a steepest descent method with an approximate gradient to minimize its loss function *stress*.

For `eulerAPE`, a good diagram is one that satisfies Equation (5.19) and is thus a wellformed 3-Venn diagram with $diagError \leq 10^{-6}$. For `venneuler`, a good diagram is one that is not necessarily a Venn diagram, even when a Venn diagram is required, with $stress \leq 10^{-6}$, so that if such a diagram is found, `venneuler` terminates the optimization before the 200 fixed number of iterations are completed. Thus, to compare the accuracy of the diagrams generated by `eulerAPE` and `venneuler`, we computed: *stress* for the diagrams generated by `eulerAPE` using `venneuler`'s version 1.1-0 source code, but `eulerAPE`'s analytic method to compute the zone areas; ***diagError*** Equation (5.18) for the diagrams generated by `venneuler` using `eulerAPE`'s source code, but `venneuler`'s numerical approximation to compute the zone areas.

Results and Discussion—Diagram Error

None of the diagrams generated by `venneuler` for the 10,000 area specifications had $stress \leq 10^{-6}$ or $diagError \leq 10^{-6}$. Thus, **none of `venneuler`'s diagrams were good** according to `venneuler`'s and `eulerAPE`'s diagram error measure. Also, only 64.5% (i.e., 6453/10,000) of the generated diagrams were 3-Venn diagrams. The other 35.5% (i.e., 3547/10,000) were 3-Euler diagrams, as one or more of the seven zones that are interior to the curves of a 3-Venn diagram were missing.

Figure 5.44A(i) and Figure 5.44B(i) are **examples of diagrams generated by `venneuler` that were not Venn diagrams**. Both diagrams had a relatively low *stress* (5.69×10^{-4} and 3.17×10^{-3} respectively), which was close to the *stress* value that a good diagram should have (i.e., $stress \leq 10^{-6}$). However, Figure 5.44A(i) was **missing zone *abc*** (despite its required area was larger than that for zones *ab* and *ac* and similar to that for zone *bc*) and Figure 5.44B(i) was **missing zone *ac*** (despite its required area was similar to that for zone *b*). Such diagrams are far more misleading than ones with inaccurate zone areas, as besides showing incorrect quantities, not all the required set relations are depicted. Figure 5.44A(i) and Figure 5.44B(i) could have been less misleading if they were depicted as the non-wellformed diagrams in Figure 5.45A and Figure 5.45B, as the missing zones could have been interpreted as non-visible because their area is incorrectly very small. Having a low *stress* for diagrams with the incorrect set of zones is a problem in `venneuler` (Section 3.7.4), making stress inappropriate for such diagrams. In contrast, the *diagError* for these diagrams was not so low (1.16×10^{-2}, 2.07×10^{-2}). Some of `venneuler`'s diagrams were also non-wellformed, impeding diagram comprehension (Section 3.7.3). Figure 5.44B(i) is non-wellformed as zone *b* is disconnected. These problems are not evident in `eulerAPE`'s diagrams (ii) and (iii) in Figure 5.44, as our diagram goodness measure Equation (5.19) and other checks during the optimization ensure that all of its generated diagrams have the required zones and are wellformed.

Figure 5.44C(i) is **an example of a diagram generated by `venneuler` that had all the required zones**. Such diagrams were often misleading, as the zone areas were inaccurate. According to the area specification for which this diagram was generated, zone *a* should be 2.1 times larger than zone *ab* and 1.7 times larger than zone *ac*. However, zone *a* in the diagram is much smaller than both zones *ab* and *ac*. The *stress* was low ($stress=4.27 \times 10^{-3}$, $diagError=2.30 \times 10^{-2}$), but greater than that of Figure 5.44A(i) and Figure 5.44B(i), despite that the latter diagrams had missing zones and were thus more misleading.
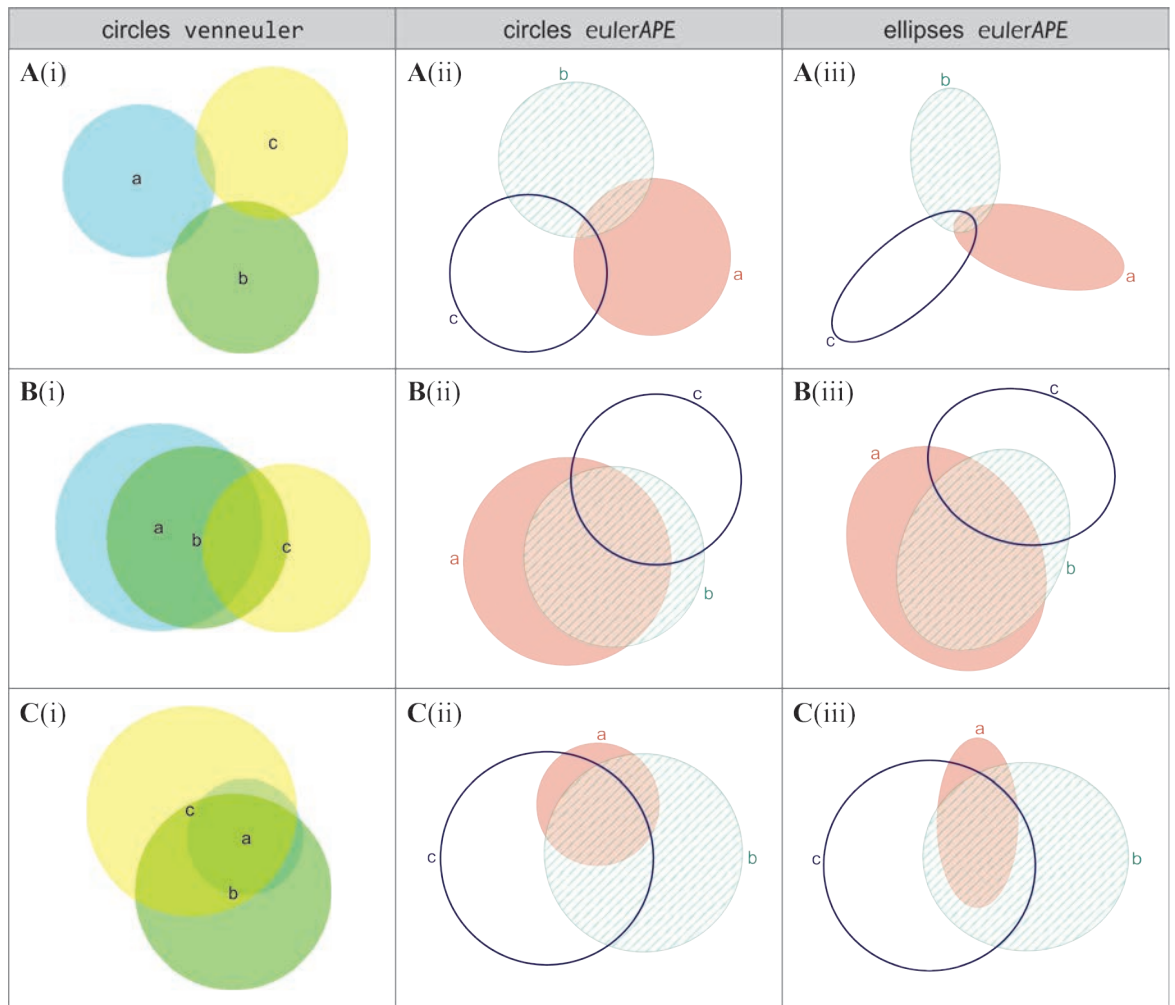
Figure 5.44: *Examples of diagrams generated with circles by* venneuler*, with circles by* eulerAPE *and with ellipses by* eulerAPE *for area specifications in* AREASPECLIB_eval. (A) For {a=3491, b=3409, c=3503, ab=120, ac=114, bc=132, abc=126}. A(i) is not a Venn diagram, as zone *abc* is missing, with $stress = 5.69 \times 10^{-4}$, $diagError = 1.16 \times 10^{-2}$. A(ii) and A(iii) are Venn diagrams. A(ii) has $stress = 8.36 \times 10^{-3}$, $diagError = 2.63 \times 10^{-2}$. A(iii) has $stress = 3.96 \times 10^{-12}$, $diagError = 6.55 \times 10^{-7}$. (B) For {$a$=45910, b=3261, c=45467, ab=58845, ac=3028, bc=16406, abc=18496}. B(i) is not a Venn diagram, as zone *ac* is missing, with $stress = 3.17 \times 10^{-3}$, $diagError = 2.07 \times 10^{-2}$. B(i) is also non-wellformed, as zone *b* is disconnected. B(ii), B(iii) are Venn diagrams. B(ii) has $stress = 2.13 \times 10^{-2}$, $diagError = 4.36 \times 10^{-2}$. B(iii) has $stress = 3.43 \times 10^{-12}$, $diagError = 6.85 \times 10^{-7}$. (C) For {$a$=3664, b=46743, c=59811, ab=1742, ac=2099, bc=17210, abc=24504}. C(i), C(ii), C(iii) are all Venn diagrams. C(i) has $stress = 4.27 \times 10^{-3}$, $diagError = 2.30 \times 10^{-2}$. C(ii) has $stress = 8.31 \times 10^{-3}$, $diagError = 2.44 \times 10^{-2}$. C(iii) has $stress = 1.13 \times 10^{-12}$, $diagError = 4.03 \times 10^{-7}$.
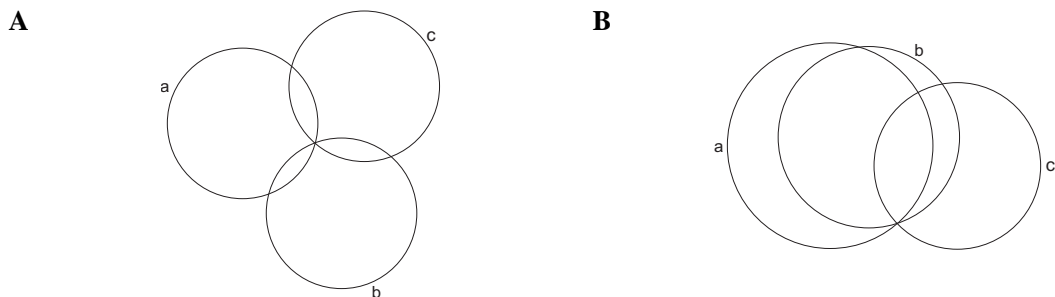
Figure 5.45: *Diagrams with circles alternative to* venneuler*'s with a missing zone, that could give the impression that the zone has a very small area.* Both diagrams have a triple point instead of: (A) the white space where zone *abc* should have been in Figure 5.44A(i); (B) the disconnected zone *b* in Figure 5.44B(i).

The diagrams (ii) generated by eulerAPE with circles in Figure 5.44 had inaccurate and misleading zone areas like those generated by venneuler, but all depicted the required set of zones and all were wellformed. All of the diagrams (iii) generated by eulerAPE with ellipses in Figure 5.44 had the required set of zones as well as $stress \leq 10^{-6}$ and $diagError \leq 10^{-6}$ and were thus considered good by both venneuler's and eulerAPE's error measure. In all of these cases, at least one elongated ellipse was required to accurately represent the area specification and thus, cannot be drawn accurately with circles.

As shown in **Figure 5.46 (for *stress*) and Figure 5.47 (for *diagError*)**, the majority of venneuler's diagrams had a lower *stress* and *diagError* than those of eulerAPE 's diagrams with circles (a lower *stress* for 8675/10,000 diagrams; a lower *diagError* for 6234/10,000 diagrams), but a greater *stress* and *diagError* than those of eulerAPE's diagrams with ellipses (a greater *stress* for 9730/10,000 diagrams; a greater *diagError* for 9660/10,000 diagrams). The range, median and mean of *stress* and *diagError* for the 10,000 diagrams by venneuler, eulerAPE with circles, eulerAPE with ellipses are provided in Table 5.2.

Table 5.2: *The stress and diagError of all the diagrams generated with circles by* venneuler, *with circles by* eulerAPE *and with ellipses by* eulerAPE, *for the 10,000 area specifications in* AREASPECLIB_eval.

	stress			diagError		
	range	median	mean	range	median	mean
circles venneuler	$[3.77 \times 10^{-5}, 6.14 \times 10^{-1}]$	3.04×10^{-2}	6.41×10^{-2}	$[1.56 \times 10^{-3}, 2.46 \times 10^{-1}]$	4.56×10^{-2}	5.73×10^{-2}
circles eulerAPE	$[1.91 \times 10^{-10}, 7.79 \times 10^{-1}]$	7.00×10^{-2}	1.13×10^{-1}	$[3.30 \times 10^{-6}, 3.31 \times 10^{-1}]$	6.28×10^{-2}	6.73×10^{-2}
ellipses eulerAPE	$[3.98 \times 10^{-14}, 2.24 \times 10^{-1}]$	7.59×10^{-12}	1.17×10^{-10}	$[6.00 \times 10^{-8}, 1.39 \times 10^{-1}]$	8.00×10^{-7}	2.94×10^{-3}

The **differences between venneuler's diagrams and eulerAPE's diagrams with ellipses** were expected due to the limitations of circles in generating accurate diagrams for most area specifications. None of venneuler's diagrams were considered good by *stress* and *diagError*, but 8529 and 8607 of the 10,000 diagrams generated with ellipses by eulerAPE were considered good by respectively *stress* and *diagError* (the difference between the percentages of good diagrams by *stress* and *diagError* for eulerAPE's diagrams with ellipse is not statistically significant—using R's pro.test with Yates' continuity correction disabled, $\chi^2(1) = 2.48$, $p = 0.12$).

The **differences between venneuler's and eulerAPE's diagrams with circles** could be less expected. A Friedman rank sum test for non-normal distributions and repeated-measure data revealed a significant effect of drawing method on *stress* ($\chi^2(1) = 5402.3$, $p < 2.2 \times 10^{-16}$) and on *diagError* ($\chi^2(1) = 609.1$, $p < 2.2 \times 10^{-16}$). Post-hoc tests using Wilcoxon tests with Bonferroni correction showed significant differences between venneuler and eulerAPE with a large effect size on *stress* ($W = 1763624$, $Z = -80.50$, $p < 2.2 \times 10^{-16}$, $r = 0.57$) and with a medium effect size on *diagError* ($W = 14730686$, $Z = -35.58$, $p < 2.2 \times 10^{-16}$, $r = 0.25$). So according to these measures, venneuler's diagrams were more accurate than those of eulerAPE. However, while all of eulerAPE's diagrams depicted the required set of zones, 35.5% of venneuler's diagrams had missing zones and yet 83.5% of these diagrams had a low stress ($stress < 10^{-2}$). So eulerAPE's diagrams could still be more helpful than those of venneuler as all and only the required set relations are depicted. Also, out of the 10,000 diagrams drawn by eulerAPE with circles, zero (i.e., 0%) had $diagError \leq 10^{-6}$ (Section 5.7.2), but 28 (i.e., 0.3%) had $stress \leq 10^{-6}$ (the difference

between these percentages is statistically significant—using R's pro.test with Yates' continuity correction disabled, $\chi^2(1) = 28.04$, $p = 1.19 \times 10^{-7}$). Thus with 99% confidence, these results indicate that **for any area specification corresponding to a 3-Venn diagram**, **a good diagram with circles and with *stress* $\leq 10^{-6}$ can be generated by**: euler***APE*** for 0.2% to 0.5%; `venneuler` for 0.0% to 0.1%.
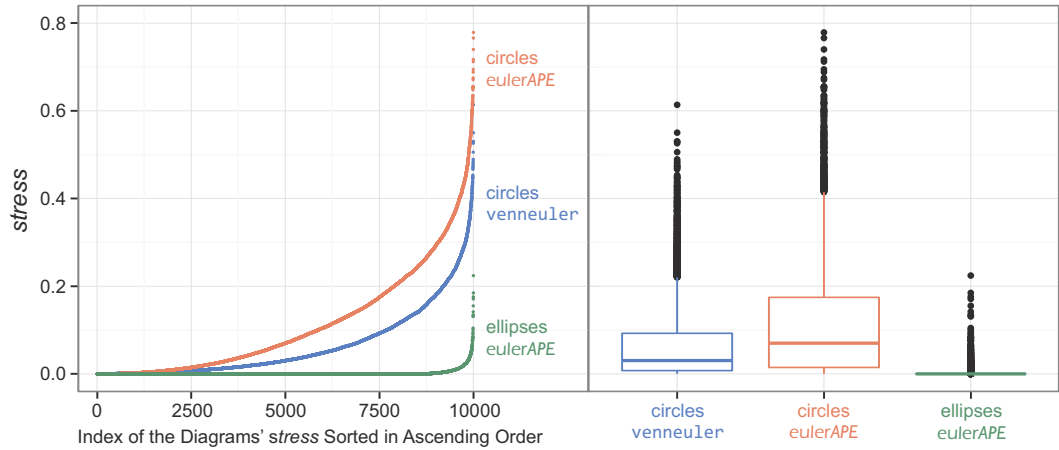
Figure 5.46: *The stress of all the diagrams generated with circles by* `venneuler`, *with circles by* euler*APE* *and with ellipses by* euler*APE*, *for the 10,000 area specifications in* AREASPECLIB_*eval*.
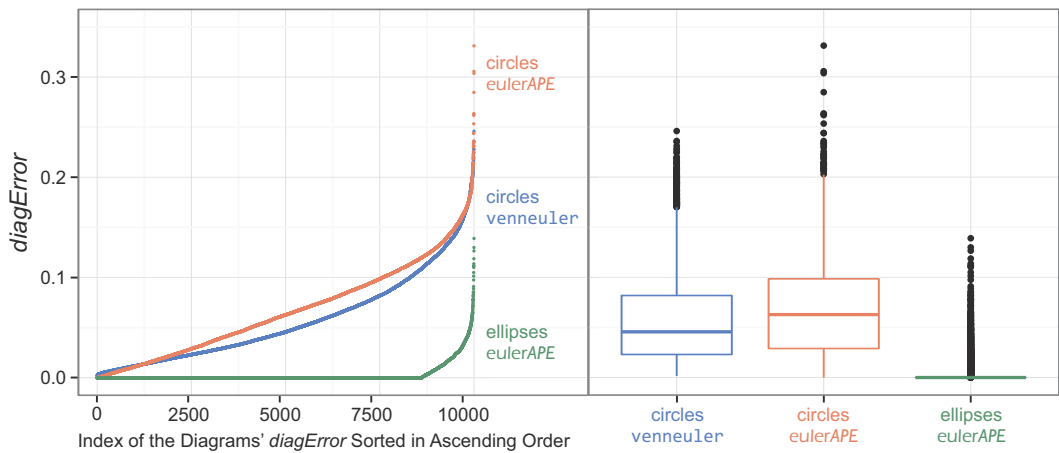
Figure 5.47: *The diagError of all the diagrams generated with circles by* `venneuler`, *with circles by* euler*APE* *and with ellipses by* euler*APE*, *for the 10,000 area specifications in* AREASPECLIB_*eval*.

Results and Discussion—Area Specifications of Close to Accurate Diagrams with Circles

According to *stress* but not *diagError*, there are a few area specifications for which an accurate area-proportional 3-Venn diagram can be drawn with circles. Out of the 10,000 diagrams with circles by euler*APE*, 28 were good according to *stress* but not *diagError* (*stress*: $[1.91 \times 10^{-10}, 9.36 \times 10^{-7}]$, median 1.40×10^{-7}, mean 2.55×10^{-7}; *diagError*: $[3.66 \times 10^{-6}, 2.68 \times 10^{-4}]$, median 9.10×10^{-5}, mean 1.04×10^{-4}). Yet the *diagError* of all of these diagrams was relatively low and close to *diagError* $\leq 10^{-6}$, which good diagrams in euler*APE* must satisfy. Figure 5.48 shows the diagrams out of these 28 with: (A) the lowest *diagError*, (B) a *diagError* close to the mean *diagError* for these 28, (C) the highest *diagError*. In these three diagrams, the depicted zone areas (available in the caption) varied by very few units from those in

the required area specification for two zones (zone *ab* and *ac* in Figure 5.48A) or more (all the zones in the case of Figure 5.48B,C). None of these diagrams are misleading and none of the inaccuracies seem noticeable. However, no empirical evidence indicates what inaccuracies in zone areas are not noticeable (Section 3.7.3), so identifying accurate diagrams using a diagram error measure and a very small value defining when error is nil ensures a more objective approach. There will always be diagrams (like those in Figure 5.48) with an error close to nil but still considered non-good, as a value for error has to be defined for the optimization to identify when a good diagram has been found. To inform the user about such cases, eulerAPE always outputs the *diagError* and the actual zone areas of the generated diagram.
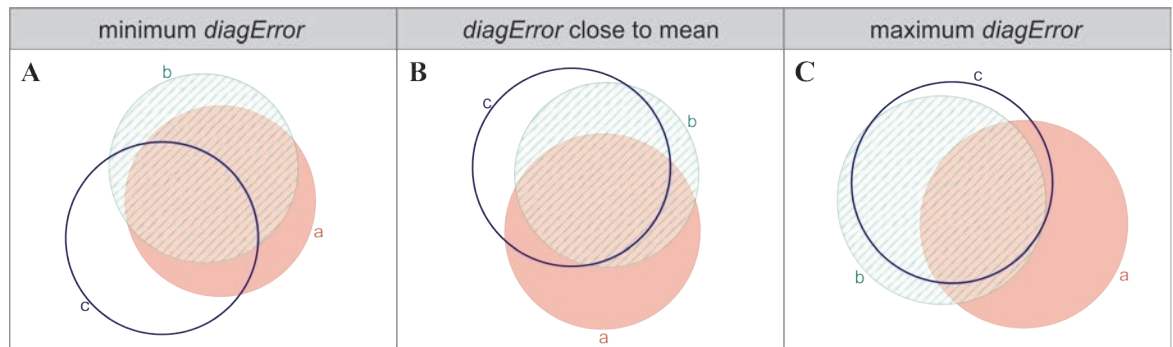
Figure 5.48: *Examples of the 28 diagrams generated with circles by* eulerAPE *for area specifications in* AREASPECLIB_*eval with stress* ≤ 10^{-6}. *(*A) The diagram with the lowest *diagError* (3.66×10^{-6}) out of the 28 diagrams, and *stress*=2.32×10^{-10} generated for {*a*=3004, *b*=4454, *c*=9513, *ab*=6956, *ac*=2785, *bc*=915, *abc*=9841}. The diagram's zone areas correspond to {*a*=3004, *b*=4454, *c*=9513, *ab*=6955, *ac*=2784, *bc*=915, *abc*=9841}. (B) A diagram with *diagError* = 9.96×10^{-5} close to the mean *diagError* (1.04×10^{-4}) for these 28 diagrams, and *stress* = 1.66×10^{-7} generated for {*a*=6631, *b*=1341, *c*=4533, *ab*=1759, *ac*=1190, *bc*=4306, *abc*=9455}. The diagram's zone areas correspond to {*a*=6629, *b*=1339, *c*=4531, *ab*=1761, *ac*=1192, *bc*=4309, *abc*=9453}. (C) The diagram with the highest *diagError* (2.68×10^{-4}) out of the 28 diagrams, and *stress* = 9.09×10^{-7} generated for {*a*=9465, *b*=2407, *c*=1156, *ab*=891, *ac*=621, *bc*=7861, *abc*=8800}. The diagram's zone areas correspond to {*a*=9469, *b*=2412, *c*=1161, *ab*=885, *ac*=616, *bc*=7853, *abc*=8806}.

All of these area specifications, for which the 28 diagrams were drawn, had a greater total area for the 1-curve zones (median 50.4%, mean 53.7%, of the diagram area) than for the 2-curve zones (median 28.1%, mean 27.4%, of diagram area), and an area for the 3-curve zone that was amongst the largest of all zone areas (the largest for 12/28 area specifications). This is evident in Table 5.3. Areas for 1-curve zones were on average twice as large as those for 2-curve zones. This contrasts the area specifications drawable with ellipses in Section 5.7.2, as the areas for the 1-curve zones were on average only 10% larger than those for the 2-curve zones. Also, the area for the 3-curve zone and the range of the zone areas were larger for these 28 area specifications than for the area specifications drawable with ellipses.

Table 5.3: *The median and mean area percentages with respect to the diagram area for the 1-curve zones, 2-curve zones, 3-curve zone and the range of zone areas, of the 28 area specifications in* AREASPECLIB_*eval for which* eulerAPE *generated a diagram with circles with stress* ≤ 10^{-6}.

	1-curve zones	2-curve zones	3-curve zone	range of the zone areas
median	16.9%	6.2%	21.9%	25.5%
mean	17.9%	9.1%	18.9%	27.3%

These observations indicate that, **if an area specification has (1) areas for the 1-curve zones that are around twice as large as those for the 2-curve zones, and (2) an area for the 3-curve zone that is larger or as large as the areas for the 1-curve zones, then it is highly likely that a close to accurate area-proportional 3-Venn diagram drawn with circles exists for that area specification.**

Results and Discussion—Time and Iterations

`venneuler` **was faster than** euler*APE*. Median and mean **generation times in seconds** were: `venneuler`, 0.6 and 0.6; euler*APE* with ellipses, 0.4 and 5.3; euler*APE* with circles, 3.2 and 3.4. This could be due to the various differences between `venneuler` and euler*APE*, such as the method used to compute the zone areas, the search space being larger and more complex for ellipses than circles, and the optimization algorithm, where example `venneuler` terminates after a maximum of 200 iterations and euler*APE* carries out up to 10 reruns to avoid local minima and generate an accurate diagram. Despite this, euler**APE** **generates more accurate diagrams than** `venneuler` **and within a time that ensure the users' attention is maintained** (Section 5.7.2). If none of euler*APE*'s reruns were considered, euler*APE*'s median and mean times in seconds were 0.4 and 0.9 for ellipses and 0.3 and 0.3 for circles, thus similar to or less than `venneuler`'s. Yet, if euler*APE*'s rerun option was disabled, good diagrams with ellipses and *diagError* $\leq 10^{-6}$ would have been drawn for 83.7% and not 86.1% of the 10,000 area specifications.

Each of `venneuler`'s diagrams were generated after 200 iterations. Median and mean **number of iterations** for euler*APE*'s 10,000 diagrams were 36 and 708 for ellipses and 500 and 521 for circles, and for only euler*APE*'s 8607 good diagrams with ellipses and *diagError* $\leq 10^{-6}$ were 35 and 41. Thus, for 86.1% of the area specifications, on average, euler*APE* required less than a fourth of `venneuler`'s number of iterations. This indicates that the high means for euler*APE*'s 10,000 diagrams were due to reruns. In fact, the median and mean number of iterations for euler*APE*'s 10,000 diagrams excluding reruns were 36 and 111 for ellipses and 49 and 51 for circles.

We continue with our comparative evaluation of the effectiveness of ellipses, circles and polygons by analysing the diagrams generated by euler*APE* and various other drawing methods for real world data.

5.7.4 Comparison with Circles and Polygons and Various Drawing Methods

Area-proportional 3-Venn diagrams are used extensively in various scientific disciplines to aid data analysis (Section 3.7.1), but often the diagrams are more misleading than helpful due to the limitations of the curve shapes used by current drawing methods (Section 5.2.1). We investigated this further using real world medical data obtained from a BMC Medicine journal article [Lenz and Fornoni, 2006].

The article discusses the results from a web-based survey that assessed whether US trainees in family medicine and internal medicine are knowledgeable of the complications, screening methods and therapy for chronic kidney disease (CKD). The selected data from this survey was comprised of sets A, B and C with participating trainees who claimed that secondary hyperparathyroidism is a complication of CKD (set A), those who screened before or at stage 3 of CKD (set B), and those who commenced therapy or referred the patient to a specialist when parathyroid hormone (PTH) reached a level of PTH >70 ng/ml (set C). The set relations and intersection cardinalities are $\omega = \{A = 0.25, B = 0.01, C = 0.11, AB = 0.10, AC = 0.29, BC = 0.03, ABC = 0.15\}$. To aid data analysis and raise awareness for better training to timely identify and manage patients with CKD, the area-proportional Venn diagram, in Figure 5.49D, with respect to the data ω was included in the article [Lenz and Fornoni, 2006].

D

A (79%)

25% 10% 1%

B (29%)

15%

3%

29%

11% C (58%)

"the proportion of respondents who indicated that secondary hyperparathyroidism is a complication of CKD (A, 79%), that they screened starting in stage 3 CKD (B, 29%), and that they intervened at an intact PTH concentration of more than 70 pg/ml (C, 58%). Overlapping areas correspond to the proportion of participants whose responses fulfilled two adjacent or all three criteria (white, 15%)."

CKD = chronic kidney disease
PTH = parathyroid hormone   (Lenz & Fornoni, 2006)

C1 STATA PVENN
diagError = 0.05

C2 Venn Diagram Plotter
diagError = 0.04

C3 3 Circle Venn
diagError = 0.14

C4 PatternLab for proteomics
diagError = 0.03

C5 BioVenn
(C) 2007-2008 Tim Hulsen
diagError = 0.04

C6 Vennerable *circles*
diagError = 0.04

C7 venneuler
diagError = 0.03

C8 Google code
missing zone BC

P1 VennMaster
missing zones C and BC

P2 Vennerable *triangles*
diagError = 0.01

P3 DrawVenn
diagError = 0.00

P4 Vennerable *squares*
diagError = 0.18

P5 Convex Venn-3
diagError = 0.00

P6 DrawEuler
diagError = 0.00

E eulerAPE
diagError = 0.00

Figure 5.49: *The Venn diagram in the journal article from where the real medical data was obtained, together with the diagrams generated by the various drawing methods for the same data.* All the diagrams are meant to depict $\omega = \{A = 0.25, B = 0.01, C = 0.11, AB = 0.10, AC = 0.29, BC = 0.03, ABC = 0.15\}$, which represents the findings of a medical survey from a journal article [Lenz and Fornoni, 2006] that also included diagram D for ω. So, D is a redrawing of the bottom diagram in Figure 5 of the article. The diagrams generated for ω using a circle-based drawing method are marked as C, those of polygon-based methods are marked as P, and the only diagram with ellipses that by eulerAPE is E. Green indicates accurate diagrams with *diagError* $\leq 10^{-6}$. Red indicates diagrams with inaccurate or missing zones.

Diagrams with respect to ω were generated using **circle-based drawing methods**, namely: **C1**, Stata's PVENN [Gong and Ostermann, 2011]; **C2**, Venn Diagram Plotter [Littlefield and Monroe, 2013]; **C3**, 3 Circle Venn [Chow and Rodgers, 2005]; **C4**, a module in PatternLab for proteomics [Carvalho et al., 2008]; **C5**, BioVenn [Hulsen et al., 2008]; **C6**, Vennerable [Swinton, 2007, 2009]; **C7**, venneuler [Wilkinson, 2012]; **C8**, Google Venn Charts [Google Inc, 2012]. Other diagrams with respect to ω were generated using **polygon-based drawing methods**, namely: **P1**, VennMaster [Kestler et al., 2005, 2008], with circle-like polygons; **P2**, Vennerable triangles [Swinton, 2009], with triangles; **P3**, DrawVenn [Chow and Ruskey, 2004], with rectilinear polygons; **P4**, Vennerable squares [Swinton, 2009], with rectangular polygons; **P5**, Convex Venn-3 [Rodgers et al., 2010a], with 4-sided and 5-sided convex polygons; **P6**, DrawEuler [Chow and Ruskey, 2005], with irregular, non-convex polygons. All the diagrams are available in Figure 5.49, together with that generated using **E**, euler*APE* **with ellipses**. The design of each diagram (e.g., labels, legend, colours, outlines, background) is precisely the same as that generated by the drawing method. Curve labels were only added to C2, P3 and P6 as no labels or legend are provided with the diagram. The numeric labels in euler*APE*'s diagram were added manually to illustrate how the diagram in the article would have looked like, if it was drawn with ellipses.

The ***diagError* of all the generated diagrams** that depicted all of the seven zones interior to the curves of a 3-Venn diagram was computed and is displayed in Figure 5.49. To calculate *diagError*, the zone areas of the diagrams drawn with circles were computed using euler*APE*'s analytic geometry method (Section 5.4.5) and the zone areas of the diagrams drawn with polygons were computed using standard geometry formula. The *diagError* was then calculated by using Equation (5.18) in Section 5.6.3 as in euler*APE*. Figure 5.49 also notes the missing zones of generated diagrams that were not Venn diagrams.

Looking into the diagrams in Figure 5.49, we note that **all the diagrams drawn with circles including D, the diagram in the article, have inaccurate zone areas and are misleading**. For instance, zone *B* (1% in ω) is much larger than zone *BC* (3% in ω), zone *C* (11% in ω) is much larger than zones *AB* (10% in ω) and *ABC* (15% in ω), and zone *A* (25% in ω) in most diagrams is larger than zone *AC* (29% in ω). These issues are also evident in C7, the diagram generated by venneuler that is considered the most effective in generating accurate diagrams using circles. Despite aiming at having a larger area for zones with a larger value, regardless of whether the zone areas are directly proportional to the values in the area specification, C3 generated by the first circle-based drawing method, 3 Circle Venn, also has zone *ABC* (15% in ω) similar in area to that of zone *AC* (29% in ω). The same is evident in D, as D was generated by the method of C3. However, C8 is the most misleading and inaccurate, as zone *BC* is missing and both zones *B* and *AB* (respectively 1% and 10% in ω) are much larger than zones *C*, *AC* and *ABC* (respectively 11%, 29% and 15% in ω). With respect to *diagError*, the most accurate were C4 and C7 (*diagError* = 0.03), followed by C2, C5, C6 (*diagError* = 0.04) and C1 (*diagError* = 0.05), and finally C3 and thus D (*diagError* = 0.14). Due to the regularity and good continuity of circles, the curves are often easily distinguishable and identifiable (Section 3.5.2). For a few (e.g., C3 and D), it is difficult to comprehend in which curves the zones are located in. However, this is down to design (e.g., colours, outlines, background), which we discuss in Section 5.8.

In contrast, **most of the diagrams with polygons are either accurate with *diagError* $\leq 10^{-6}$**, as P3, P5, P6, **or have zone areas that are less misleading than those of diagrams with circles**, as P2, P4. The latter is true as, for instance, consistent with ω, zone *B* is always the smallest and zone *AC* is always the largest. The only diagram with missing zones (zones *C* and *BC*) is P1. The drawing method used to generate P1, VennMaster, is non-deterministic and so this is one of the possible diagrams

generated for ω. Since the curves are depicted as regular, circle-like polygons, this drawing method has the same limitations as others that use circles. Though the other diagrams with polygons are more accurate than those with circles, the curves lack good continuity and are not easily identified (Section 3.5.2). Curve identification is particularly difficult when the curves meet at bending points, as P5 and P6, and when the curves are non-convex, as P6. Moreover, all of the diagrams, except for P1, are non-wellformed, making these diagrams accurate but difficult to use (Section 3.7.3). Thus, the preference for the less accurate diagrams with circles instead of polygons.

Using ellipses, diagram E has zone areas that are accurate and directly proportional to the quantities in ω ($diagError \leq 10^{-6}$). It is also easy to comprehend and to analyse the depicted data, as the curves are regular and have good continuity like circles. So ellipses can be more effective than both circles and polygons in drawing accurate, easy to comprehend area-proportional 3-Venn diagrams. This was also demonstrated with other real data in Section 5.2.2, where Figure 5.3 and Figure 5.4 illustrate the accurate diagrams generated by eulerAPE with ellipses as alternatives to respectively the misleading diagrams drawn with circles in Figure 5.1 and the incomprehensible diagrams drawn with polygons in Figure 5.2. Being the only method that uses ellipses, the effectiveness of ellipses could be the primary reason why eulerAPE is being used in various areas and why its diagrams are appearing in various journal articles (details in Section 5.1). The design of eulerAPE's diagrams is also different from that of other diagrams. As we will discuss in Section 5.8, eulerAPE's design aids curve and zone identification.

Another area-proportional 3-Venn diagram for the same data sets but for the management of the anaemia rather than secondary hyperparathyroidism (so set *C* was based on a level of hemoglobin rather than parathyroid hormone) was included in the article as shown in Figure 5.50. Multiple Venn diagrams depicting the same data sets but for a different factor are commonly used to easy note differences. However, the zone areas have to be directly proportional to the required data for them to be helpful.

As explained earlier, Figure 5.50A(ii) is misleading due to inaccuracies in the zone areas. **Figure 5.50A(i) could be considered more misleading** as besides inaccuracies in the zone areas (e.g., zone *B* with label 3% is larger than zone *AC* with label 4%; zone *A* with label 36% is larger than zone *AB* with label 41%), a 3-Venn diagram is depicted when the data {*A* = 0.36, *B* = 0.03, *C* = 0.00, *AB* = 0.41, *AC* = 0.04, *BC* = 0.00, *ABC* = 0.11} corresponds to a 3-Euler diagram with no zones *C* and *BC*. So the depicted set relations are not entirely correct. Currently eulerAPE draws highly accurate 3-Venn diagrams even when zones are very small and barely visible. So, as discussed in Section 5.7.1, we could use eulerAPE to generate an accurate diagram with respect to {*A* = 0.36, *B* = 0.03, *C* = 0.00001, *AB* = 0.41, *AC* = 0.04, *BC* = 0.00001, *ABC* = 0.11}. As illustrated in Figure 5.50B(i), the minor differences between the area of zones *C* and *BC* of eulerAPE's diagram and those required by the actual data are not noticeable. Looking at this and the accurate Figure 5.50B(ii), it is easily noticeable that for instance, though most participants claimed that anaemia and secondary hyperparathyroidism were complications of CKD (set A), timely screening (set B) was more common with anaemia (i) than secondary hyperparathyroidism (ii), while commencement of therapy or referral to a specialist (set C) was more often delayed when diagnosis was based on the hemoglobin level (i) than on the parathyroid hormone level (ii). The article emphasizes the importance for practitioners to have the right skills to manage CKD, and using the diagrams in Figure 5.50A, they attempted to raise awareness of the need to provide better training to avoid delays in screening and therapy. However, as discussed, this could have been done more effectively if their diagrams accurately depicted their findings using ellipses.
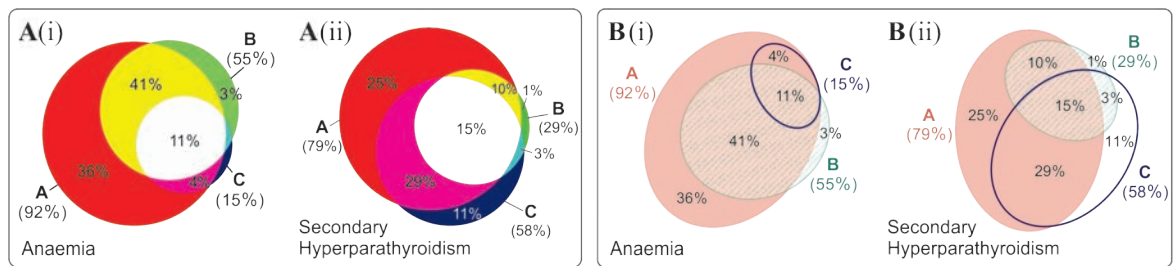
Figure 5.50: *The figure in the journal article with two Venn diagrams with circles (left, A) and the figure as it would have looked like if the diagrams were generated using* eulerAPE *and ellipses (right, B).* Figure 5 of the article [Lenz and Fornoni, 2006] was redrawn and the two Venn diagrams were placed next to each other rather than one below the other. The labels for eulerAPE's diagrams were added manually.

We will now discuss diagram design features that could aid comprehension of the depicted set relations and set intersection cardinalities. Most of these features have already been adopted by eulerAPE.

5.8 Diagram Design Features to Aid Comprehension

To facilitate comprehension, an area-proportional Venn or Euler diagram should have curves that are easily identified and zone areas that are easily comparable to one another [Chow and Ruskey, 2004]. We discuss design features that could facilitate both.

5.8.1 Curve and Zone Identification

As demonstrated in Section 5.7.4 and discussed in Section 3.5, **regular, convex curves that have good continuity facilitate curve identification**. So curves represented as circles are preferred over polygons. However, design features of the diagram may also facilitate or hinder curve and thus zone identification.

In Figure 5.49, C2, C3, and C6 **use circles**. Yet it is unclear which zones are in which curves, as completely unrelated colours are used for the zones in each curve. Also, the various colours used hinder the visual search for a specific hue [Ware, 2008, p. 74] and thus a specific zone. It is even more difficult in C3 than C2 and C6, as the curves in C3 do not have an outline and the diagram has low luminance contrast among zones (e.g., zones *A*, *AC*) and between the zones and black background (e.g., zone *B*).

The main objective of Venn and Euler diagrams is to depict overlaps and relationships between the sets and so, the main focus has to be on the curves representing the sets rather than the individual zones representing the intersections. The zones are meaningful only when **the curves are visually distinct and the curves in which the zones are located can be identified**.

Using a **colour per curve** (for the interior) and transparency as in C4, C5, C7 and C8, the diagram is typically perceived as three overlapping curves, thus facilitating the identification of the different curves and the zones in each one. However, since the colours of the curves could perceptually fuse at overlaps, the colours of the zones in the same curve could seem unrelated, such that the transparency effect is not perceived. In fact, for transparency to be perceived, the colour values of the zones in each curve must be within a specific ratio [Masin, 1997] and a few perceptual rules have to be satisfied [Metelli, 1974]. If the interior of the curves is not coloured and the curves only have an outline, then based on the law of Prägnanz [Koffka, 1935] (stating that the simplest forms are perceived), the diagram is still perceived as

three circles rather than seven irregular shapes. However, identifying the curves in which a zone is located is not easy, even if coloured outlines as in C1 are used for the curves.

To ensure that the curves are visually distinct, different **visual feature channels** should be used [Ware, 2012, p. 145 G5.2]. This is also known as a heterogeneous channel-based approach [Ware, 2008, p. 52]. At the early stages of visual processing, different feature types, such as colour and texture, are processed separately and in parallel by different cells and each type has its own channel such that information in one channel, example colour, does not interfere with another, example texture [Ware, 2012]. When colour is the only feature type used to distinguish the curves, the same and only channel is used to retrieve information on different curves and so, unless specific colour value ratios and other rules are satisfied, transparency is not perceived and the layers perceptually fuse (discussed earlier). To avoid this problem, euler*APE* uses colour, texture and outline to ensure the curves are visually distinct (e.g., Figure 5.49E). Since these features are processed separately and are expressed in three different channels, three separate layers one per curve are perceived and none of them perceptually fuse. So, the curves in which the zones are located are easily identified. If viewers tune their attention to the feature type of a specific curve, other feature types recede [Ware, 2008], thus facilitating reasoning about the sets. In fact, Ware advices: "To define multiple overlapping regions, consider using a combination of line contour, color, texture, and Cornsweet contours" [Ware, 2012, p. 188 G6.6].

Even so, euler*APE* still uses **distinct curve colours**. Unique hues, including colours black, white, red, green, yellow and blue [Berlin and Kay, 1969], are distinct and contrasting [Mullen, 1985], processed in parallel [Hurvich, 1981], and primary colours understood across cultures [Berlin and Kay, 1969]. Thus, the colours used by euler*APE* are red, green and blue. Though colour-blind readers would have difficulties distinguishing red and green, they would still be able to distinguish the curves, as one is solid filled and the other is textured filled. In fact, the curves would still be visually distinct, even if the diagram is displayed black and white. Blue-yellow colour blindness is less common. However, if these colours were used instead of red and green, choosing a background that ensures adequate luminance contrast for both colours would be difficult, as yellow requires a black background (e.g., curve *B* in Figure 5.49 V7 is not easily visible on white), while blue requires a white background (e.g., zone *C* in Figure 5.49 V3 is not easily visible on black) [Ware, 2008, p. 70]. This is not a problem for red and green as a white background ensures adequate luminance contrast for both colours.

5.8.2 Zone Area Judgement and Comparison

Aesthetic characteristics, such as the shape of the zones and the curves, could facilitate or hinder zone or curve area comparison, but no empirical studies have been carried out for such diagrams (Section 3.7.3). Humans are biased to area judgement [Cleveland and McGill, 1984] and so Cleveland [1985, p. 282] argues that quantitative data should be depicted using position or length which humans are able to judge more accurately than area. However, Venn and Euler diagrams are widely used (Section 3.7.1). In Section 3.7.3, we discussed ways to counteract judgement bias.

For instance, **an additional plot that uses position or length judgement** could be provided with the diagram, as in Figure 5.51, which was generated as one figure by a prototype variant of euler*APE* that we have devised (but not yet released) and which illustrates Figure 5.49E but with an adjacent bar chart as an alternative representation of the quantities depicted by the zones in the Venn diagram. With our design for the curves, it is possible to relate to the zones in the Venn diagram by only showing a sample

of the interior fill of the zones in the bar chart without requiring any labels. This facilitates the mapping of the plot to the diagram and reduces the processing time as the visualization is seen as one rather than two. Also, numeric labels in the zones should be avoided, to prevent data redundancy, which could delay the processing of the visualization [Tufte, 1990], and to instigate the use of the bar chart whenever the quantities depicted by the zones are compared and judged, such that any biases are avoided. However, the effectiveness of such visualizations needs to be evaluated.
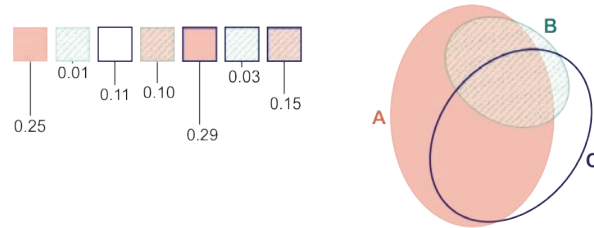
Figure 5.51: *A hybrid diagram generated by a prototype variant of* eulerAPE *to facilitate zone area judgement and comparison.* The Venn diagram is the same as that in Figure 5.49E.

Alternatively, **equally-sized glyphs** that are proportional in number to the quantities in the area specification could be placed inside the zones of a Venn or Euler diagram that is not necessarily area-proportional. This would replace area judgement with cardinality judgement of a number of discrete objects. If the glyphs are arranged in multiple blocks with a size that is within a subitizing range (i.e., the number of objects that humans can effortlessly and correctly judge their number without counting), the number of glyphs is easily and accurately estimated without the need for focused attention [Cowan, 2000; Miller, 1956]. Euler diagrams with glyphs were discussed in Section 3.8 and for Bayesian reasoning in Section 3.9.2. In Chapter 6, we discuss our drawing method eulerGlyphs that draws Euler diagrams with glyphs for Bayesian problems and a study we conducted to assess the effectiveness of these diagrams.

5.9 Future Work

This is the first work to assess the effectiveness of ellipses in drawing accurate diagrams with smooth curves to depict the required set relations and cardinalities. Our novel drawing method, eulerAPE, is also the first to use ellipses. Being the first, this work focused on 3-Venn diagrams. We have shown that accurate 3-Venn diagrams with ellipses can be drawn for a large majority of random area specifications (Section 5.7.2), far more than is possible with the circles, which are highly preferred over polygons.

These results indicate great potential for using ellipses to draw area-proportional 3-Euler diagrams and possibly other Euler diagrams with more than three curves. The method adopted by eulerAPE already draws some Euler diagrams, but further evaluation is required to identify abstract descriptions for which an Euler diagram can be drawn with ellipses and ensure that eulerAPE draws such diagrams adequately. Area specifications for which an accurate area-proportional 3-Euler diagram can be drawn with ellipses should be identified. This work should then be extended for diagrams with more than three curves.

Based on the current automatic drawing methods, if an accurate diagram cannot be drawn with ellipses, then polygons have to be used. However, there are other shapes that are more general than

ellipses and similarly have good continuity (e.g., ovals) that could be used. Thus, other shape representations for the curves should be investigated, so that, before a diagram is drawn, different shaped curves are considered in a "natural progression" [Chow, 2007, p. 83], starting off with circles and progressing with more general shapes such as ellipses and ovals before less desired and non-smooth shapes such as convex and non-convex polygons are considered. This ensures that the smoothest and most regular curves possible for the required area specification are used.

In our evaluation, characteristics of random area specifications (made up of random quantities) for which an area-proportional 3-Venn diagram was drawn accurately with ellipses using eulerAPE were identified (Section 5.7.2). Such observations will aid in the formalization of analytic methods that will determine whether an accurate area-proportional diagram can be drawn with ellipses for a given area specification. If similar analytic methods are formalized for other shaped curves, the most regular and smoothest curves that are appropriate for a given area specification could be identified immediately.

Besides the shape of the curves, diagram design features (e.g., colours) can also facilitate or impede understanding of the diagram and the depicted data. A few features have been discussed in Section 5.8. However, various others should be evaluated to aid curve and zone identification, facilitate reasoning about the depicted quantitative data and avoid area judgement biases. Features that could aid understanding for different users with different spatial and numeracy abilities, as well as the benefits of interactive diagrams in assisting the users with their data analysis and reasoning about the sets should also be investigated. Different labelling strategies for the curves and zones should be evaluated to assess whether they aid or hinder reasoning and area judgement and whether they add noise to the diagram.

A number of the studies should be conducted to understanding how such diagrams are processed perceptually and cognitively, how zone areas are perceived, the effect of the shape of the zones and curves on area judgement, what discrepancies in areas are not noticeable and whether perceptual scaling measures like those proposed for map symbols in cartography [Montello, 2002] (highly criticized by Tufte [1983]) aid or hinder area judgement in these diagrams. In this work, a good and accurate diagram was one whose zone areas were directly and accurately proportional to the quantities in the area specification. Doing so, it was possible to precisely assess for which area specifications an accurate, good diagram can be drawn with ellipses. However, as noted in our evaluation (Section 5.7.1, Section 5.7.2), minor errors in the zones are often not noticeable. So, the findings of these studies will aid in the identification of diagrams whose zone area errors are not human detectable. In this way, an inaccurate diagram drawn with ellipses for an area specification could be considered accurate for human use and perception and so, it could be drawn with ellipses rather than other more complex curves with less desirable features, such as irregular and non-smooth polygons. A study should also determine if numeric labels in the zones could conceal errors in zone areas, thus allowing the use of smooth curves.

Following these studies, aesthetic criteria, metrics and cognitive measures as well as perceptual and design guidelines defining an effective, good diagram for human use that facilitates comprehension and reasoning about the depicted data should be formalized and prioritized. A variant of eulerAPE should then be devised to optimize such measures, such that a diagram that is the best compromise between zone area accuracy and aesthetics is generated. Such a diagram should ideally have all the important aesthetic features and none of its zone area inaccuracies should be noticeable to the human user. This would be particularly important for those area specifications for which an accurate diagram with specific aesthetic features cannot be drawn (e.g., area specifications for which an accurate area-proportional wellformed

3-Venn digram cannot be drawn using convex curves Section 3.7.5). With such a compromise, an inaccurate diagram with smooth curves whose errors are not human detectable could be generated. Other future considerations regarding the aesthetics of such diagrams are discussed in Section 3.5.4.

It is also interesting to assess the effectiveness of allowing users to select aspects of the diagram which they consider important and would like to be optimized. Such aspects could include aesthetic features as the wellformedness properties or the shape of certain curves, as well as the accuracy of all or specific zones. The latter could aid in cases when specific zones are of greater importance.

5.10 Summary

In various application areas, area-proportional Venn and Euler diagrams, particularly 3-Venn diagrams, are used to easily visualize both set relations and cardinalities. Current automatic drawing techniques for area-proportional 3-Venn diagrams either use circles or polygons to represent the data set curves. Circles are often preferred over polygons due to their regularity and smoothness. A few use polygons to ensure the generation of accurate area-proportional diagrams for more area specifications than is possible with circles. In this chapter, we have demonstrated that using ellipses and our novel drawing method eulerAPE, accurate area-proportional 3-Venn diagrams can be drawn for a large majority of random area specifications (86.1%, $N = 10000$, 99% confidence interval 85.2% to 86.9%). So curves that are smooth like circles but more general like ellipses should be considered whenever a diagram cannot be drawn accurately with circles. This result opens a wider research question as to whether curves with different degrees of freedom such as circles, ellipses, ovals, n-ellipses, regular n-gons and irregular n-gons could be progressively considered from the most specific to the more general until a curve type that generates an accurate diagram for the required area specification is found.

Humans are biased to area judgement [Cleveland, 1994], but it is still unclear how areas of regions in an area-proportional diagram are perceived and judged and what discrepancy in the required and actual zone areas is not noticeable (discussed in Section 3.7.3). Some perceptual scaling methods have been proposed for proportional symbols on maps [Montello, 2002] (details in Section 3.7.3), but they are ineffective and non-universal as they depend on various variables such as the reader's perceptual abilities in judging area, experience, the context, and the object being judged. They could thus convey the wrong quantitative data in different contexts and to readers who are capable of perceiving areas correctly. So, if area is used to depict quantitative data then, the represented area "should be directly proportional to the numerical quantities represented" [Tufte, 1983, p. 56]. A better alternative is to avoid areas altogether [Cleveland, 1994] and instead use more effective representations such as grids of equally-sized glyphs (i.e., frequency grids).

The first and foremost use of Venn and Euler diagrams is to represent relationships amongst sets. So, a not area-proportional Venn or Euler diagram with equally-sized glyphs placed in the minimal regions of its zones could be more effective than an area-proportional diagram, such that depicting quantitative data using a number of equally-sized discrete objects is more effective than using area? We investigate this further in the next chapter for probabilistic judgement and Bayesian reasoning, which areas can use Euler diagrams to clarify the set relations of the problem.

Chapter 6

Assessing the Effect of Euler Diagrams with Glyphs on Bayesian Reasoning

This chapter outlines a user study that assessed the effect of area-proportional Euler diagrams, glyph representations and Euler diagrams with glyphs on Bayesian reasoning. It introduces euler*Glyphs (*`http://www.eulerdiagrams.org/eulerGlyphs`*) as an automatic drawing method to generate such diagrams. The study also suggests ways how visualizations can be combined with textual information.*

6.1 Introduction

Area-proportional Venn and Euler diagrams are used in various areas to depict set relations and their cardinalities (Section 3.7.1). However, humans are biased to area judgement [Cleveland, 1994] and it is still unclear how region areas are perceived and what differences in areas are not noticeable. Scaling mechanisms for proportional symbols on maps have been proposed, but are non-universal. Thus, the area should either be "directly proportional to the numerical quantities represented" [Tufte, 1983, p. 56] or avoided altogether, as discussed in Section 3.7.3 subsection 'Counteracting Area Judgement Bias'.

Grids of equally-sized glyphs, also known as frequency grids, have been used extensively in areas such as risk communication to visualize quantities and facilitate reasoning (e.g., [Zikmund-Fisher et al., 2008a, b]; Section 2.5.5). In probabilistic judgement and Bayesian reasoning, psychology studies suggest that frequency grids facilitate logical reasoning by depicting the quantities in the problem as discrete objects (Section 2.5.5), while Euler diagrams clarify the set relations and how the quantities relate (Section 3.9). Both the quantities and the set relations are important for such statistical problems. Thus, combining the two into one diagram, so that equally-sized glyphs that are proportional in number to the quantities in the problem are enclosed by the curves of an Euler diagram that only depicts the data set relations (as in Section 3.8), could be a good alternative to an area-proportional Euler diagram. Bayesian reasoning is affected by cognitive judgemental errors and biases (Section 2.5.2) and so, depicting quantities using glyphs rather than area could avoid additional biases elicited by area judgement.

In this chapter, we investigate whether an Euler diagram that is not area-proportional but which has equally-sized glyphs that are proportional in number to the quantitative data can be more effective than an area-proportional Euler diagram for Bayesian reasoning. We discuss a study we conducted to assess the effect of a textual and six visual representations including area-proportional Euler diagrams, frequency grids and the Euler diagrams with glyphs for three classic Bayesian problems. We introduce the novel method euler*Glyphs* (`http://www.eulerdiagrams.org/eulerGlyphs`) which we devised to automatically draw the visualizations investigated in the study for Bayesian problems. Laymen and professionals have difficulty in understanding statistical information and Bayesian problems, and are unaware of their wrong judgements. This could have severe consequences such as overdiagnosis of diseases (Section 2.5.1). Thus considering such an application area was a priority. This study is also important in indicating for the first time the effectiveness of different visualizations for Bayesian reasoning on a large, diverse group of participants through crowdsourcing and in suggesting novel ways how visualizations should be combined with textual information.

We will start by outlining the objectives of this work (Section 6.2), followed by our diagram drawing method euler*Glyphs* (Section 6.3), the rationale of our study design (Section 6.4), a detailed account of our two experiments (Section 6.5, Section 6.6), including our findings and their implications, future work (Section 6.7) and a summary of our contributions and this chapter (Section 6.8).

The research work in this chapter was funded by INRIA, France. It was carried out during my internship with the AVIZ team and advisors Dr Pierre Dragicevic and Prof Jean-Daniel Fekete. This work has been published in a journal [Micallef et al., 2012] and presented at IEEE InfoVis 2012 as part of IEEE VIS (formerly VisWeek) 2012, where it received a Best Paper Honourable Mention Award.

The experimental stimuli and data of the study, as well as, a video with a visual explanation of a classic Bayesian problem using Euler diagrams and glyphs are available at `http://www.aviz.fr/bayes`.

6.2 Motivation and Objectives

Let us reconsider the classic mammography Bayesian problem:

> The probability that a woman at age 40 has breast cancer is 1%. According to the literature, the probability that the disease is detected by a mammography is 80%. The probability that the test misdetects the disease although the patient does not have it is 9.6%.
>
> If a woman at age 40 is tested as positive, what is the probability that she indeed has breast cancer?
>
> [Eddy, 1982]

As mentioned in Section 2.5.2, the answer to this problem is only 7.8%. In real-life, a good approximation rather than a correct answer is sufficient. However, in a study [Eddy, 1982], 95 of 100 doctors said that the answer is between 70% to 80%. So, why is the answer so low?

The answer can be computed using Bayes' theorem (as explained in Appendix C.2.1), but this is only possible if the reader is aware of the theorem and knows how to apply it to the Bayesian problem. An alternative way is to use visualizations.

We know that 1% of the women aged 40 have breast cancer.

So, if the dots represent women, then the red dots are women with cancer and the blue dots are women without cancer.
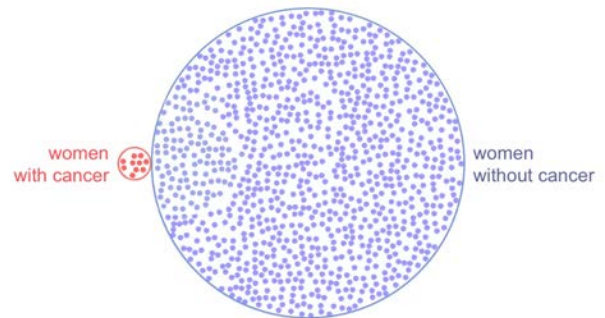
The women take a mammography.

80% of the women with cancer get a **positive mammography**
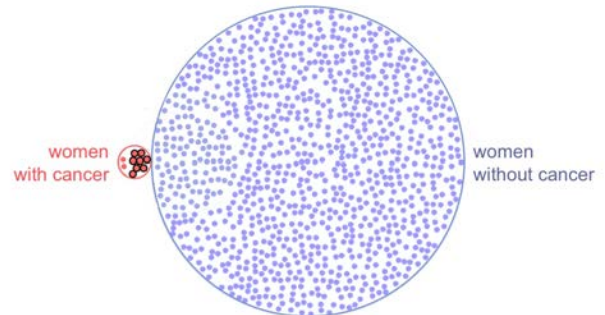
Similarly, 9.6% of the women without cancer also get a **positive mammography**.
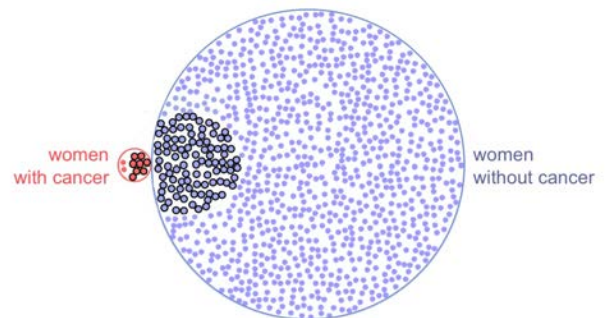
Barbara got a **positive mammography**, so she must be one of the women in the **black group**.

Thus, the probability that she is a red dot within the **black group** and that she indeed has breast cancer is only 7.8%.
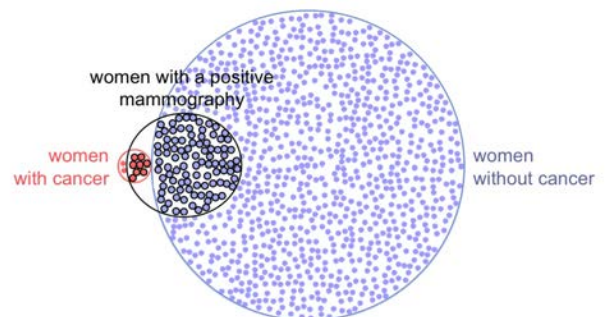
A video with this visual explanation is available at `http://www.youtube.com/watch?v=D8VZqxcu0I0`. The visualization was generated using our drawing method *eulerGlyphs*, discussed in Section 6.3.

The main problem with Bayesian inferencing is that often the base rate is ignored, leading to the base-rate fallacy (Section 2.5.2). It is unclear whether this fallacy could be avoided using visualizations that make the base rate and the nested-set relations of the problems easily visible and understandable.

6.2.1 Can Visualizations Facilitate Bayesian Reasoning?

Studies (e.g., [Sedlmeier and Gigerenzer, 2001]) indicate that learning Bayesian reasoning using visualizations helps, despite that not all users have abilities to handle graphical information [Gaissmaier et al., 2012]. A few visual representations have been proposed in psychology but most are difficult to

understand and need training prior to use (Section 2.5.4). In contrast, Euler diagrams and frequency grids are easy to understand and are widely used. Psychology studies suggest that while Euler diagrams clarify the nested-set relations of Bayesian problems (Section 3.9.1), frequency grids facilitate logical reasoning (Section 2.5.5). So combining the two into a hybrid diagram as in the previous section seems beneficial (Section 3.9.2). Also because, people with high spatial abilities are better at handling visualizations with spatial properties and those with high numeracy abilities are better at handling discrete countable objects or glyphs [Kozhevnikov et al., 2002, 2005]. The findings of previous studies are somehow inconsistent and it is still unclear which is the most effective representation for Bayesian reasoning (Section 2.5.5, Section 3.9). This could be due to the limitations of the studies.

6.2.2 Limitations of Previous Studies

Diagram Designs

Some diagram designs were inconsistent and inappropriate. For instance, the Euler diagram used in Sloman et al.'s [2003] study, shown in Figure 6.1A, was meant to be area-proportional with respect to the quantitative data of the problem (i.e., 1,000 in 'All People'; 1 in 'Number of people who have the disease'; 50 in 'Number of people that test positive for the disease'). However, an accurate area-proportional Euler diagram with respect to the problem's data would look like Figure 6.1B. Thus, Figure 6.1A could be misleading as it conveys the wrong information. Similar problems are evident in the diagrams of other studies such as Brase's [2009] Euler diagram with glyphs where the number of glyphs is not proportional to the problem's quantitative data (Section 3.9.2).
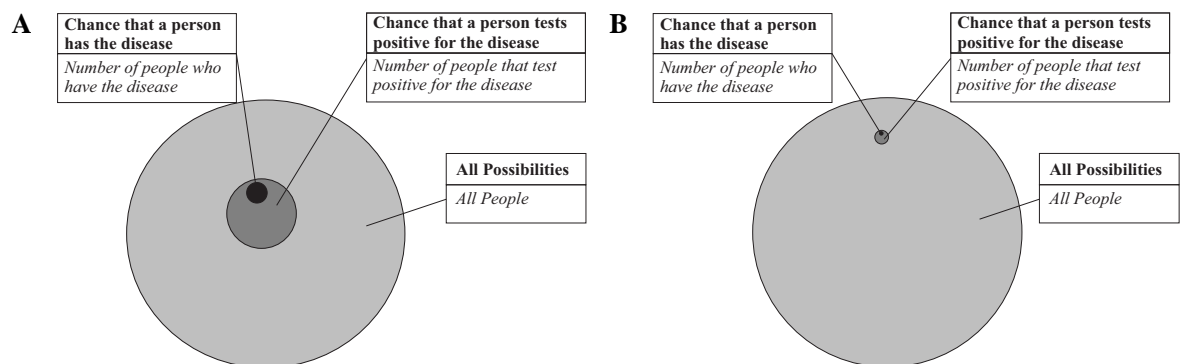
Figure 6.1: *The area-proportional Euler diagram in Sloman et al.'s [2003] study (left, A) is inaccurate with respect to the quantitative data of the tested Bayesian problem.* (A) Sloman et al.'s [2003] diagram for a disease diagnosis problem with 1,000 for 'All People', 1 for 'people who have the disease', 50 for 'people that test positive for the disease'. *Source:* [Sloman et al., 2003]–Figure 1 (B) An accurate area-proportional Euler diagram, which we manually drew, for Sloman et al.'s [2003] problem with the same quantitative data.

The rationale for the design of the evaluated diagrams is often not explained, despite that various design features could hinder the diagram's effectiveness (e.g., [Ancker et al., 2006; Tufte, 2001; Ware, 2012]; Appendix C.2.8). The diagrams were manually drawn and no automatic drawing method was devised to accurately generate the visualizations. Few diagram designs were evaluated in the individual studies, and no study assessed both visual and textual representations (Section 2.5 and Section 3.9).

Participants

Most of the studies were carried out on populations with a specific background, typically university students with around 18 to 24 years of age (Section 2.5.5, Section 3.9). In some studies, the students were from psychology and carried out the experiment to partially fulfil a course requirement (e.g., [Brase, 2009; Cole and Davidson, 1989]). In others, the students were from highly ranked universities (e.g., Stanford University [Cosmides and Tooby, 1996]). So the participants of these studies: had certain skills and cognitive abilities and knew about Bayes' theorem; were highly-focused; have experience in carrying out tests and possibly other research experiments; were aware that this was an experiment and were thus less attached to the problem; had an incentive be it monetary or to fulfil a course requirement or help a colleague or professor to complete their research (details in Section 2.5.5 and Section 3.9).

Also, human cognitive capabilities to process information decline with age, with best performance being in early twenties [Salthouse, 1996]. Similarly, the capacity of the human working memory and spatial processing abilities decrease with age [Hale et al., 2007]. So university students (typically aged 18 to 24 years) have greater cognitive abilities and a working memory with a greater capacity than other older age groups, thus they are less likely to make mistakes than the rest of laypeople [Kellen, 2012]. Individual differences in working memory are also linked to intelligence [Engle et al., 1999].

Thus the findings of most studies are not generalizable to the more diverse population of laypeople with various backgrounds, age groups and cognitive abilities. Another issue is ecological validity, since probabilistic reasoning in the real-world is different from a university setting where students are trained to remain focused and make the best of their cognitive resources to solve a problem given by a tutor.

We know of only two studies on Bayesian reasoning (assessing Euler diagrams, frequency grids or combinations of both) that were carried out on a more diverse population. Both were still carried out in a laboratory. These include the study by Kellen [2012] who had students and staff from different faculties at the University of Kentucky with 18 to 67 years of age (one third were over 22 years of age and some did not even attend high school) and that by Garcia-Retamero and Hoffrage [2013] who had doctors and patients with 18 to 85 years of age from four different hospitals in Spain. Both had an incentive: Kellen's participants took part in a lottery; Garcia-Retamero and Hoffrage's participants were paid. However, their findings might not be fully generalizable. Kellen's experiment was carried out in an academic research environment, so participants could have been subject to experimental biases like demand characteristics [Nichols and Maner, 2008]. Also, more than half (68%) had "Probability Experience" and many (78%) were completing an undergraduate or higher degree. Garcia-Retamero and Hoffrage's participants were doctors and patients and their problems related to disease diagnosis.

Difficulties in Replicating Previous Findings

Differences in "participant-sampling methodologies" [Brase et al., 2006] make it difficult for a study to replicate previous findings. A reason why the findings of previous studies were inconsistent. In fact, a series of experiments [Brase et al., 2006] demonstrate that studies using undergraduate students from top-tier national universities who are also paid are likely to claim higher success rates (e.g., Cosmides and Tooby's [1996] study whose participants were Stanford University paid students). The experiments also demonstrated that when the students are either from a second-tier regional university or unpaid, success rates drop significantly, particularly when they are not following an Honours degree course. The

authors advise "to advocate an approach that attends to the relative levels of performances across different conditions, within the same participant populations and using the same experimental methodology" [Brase et al., 2006, p. 974]. Thus the need to evaluate more diagram designs in one study.

Bayesian Problems and Their Representation

Various studies involved one Bayesian problem relating the diagnosis of a "Disease X" or "a disease" (Section 2.5.5, Section 3.9). This makes the findings less generalizable to different contexts. Also, a story about a "Disease X" is not realistic and could be difficult to relate to it. Some could have lost trust in the presented information and story (trust is essential to effectively convey a story [Glassner, 2004]). Participants may have different backgrounds, interests and skills, all of which may affect their understanding of and adherence to the problem [Brown et al., 2011; Davis et al., 2001]. A study over diverse Bayesian problems can help balance out these effects and ensures findings are generalizable.

Also, for the results of the different studies to be comparable and for possible improvements to be declared, a few of the problems including the text and quantitative data should be the same, as minor differences could have a different effect [Girotto and Gonzalez, 2001]. This also applies to the way the text and the visualization are presented together. For instance, while most studies compare text alone with exactly the same text accompanied by their visualization (e.g., [Brase, 2009; Cosmides and Tooby, 1996; Sloman et al., 2003]), Kellen [2012] removed all the quantities from the text and placed them in the visualization (as in Figure 3.37). So an improvement in Kellen's results does not necessarily mean that his visualizations are more effective. Instead, the improvement could be attributed to the fact that the participants were forced to link the diagram to the text, as the quantities were only in the diagram.

Quantitative Data of Bayesian Problems

Different quantities assigned to the base rate, hit rate and false alarm rate of the tested problems may have a different effect on the participants' accuracy and processing time [Verguts et al., 2005] due to the distance effect (numbers that differ by a large quantity are easier to compare than ones that differ by a small quantity) and the size effect (numbers that are smaller in magnitude are easier to compare than others that differ by the same quantity but are larger in magnitude) [Moyer and Landauer, 1967]. Also, base rates of a certain magnitude can mislead subjects into focusing on some values underestimating the rest [Cole, 1989]. A study over problems with values varying in distance and size could counterbalance such effects. Yet only few studies consider this factor [Cole and Davidson, 1989; Kellen, 2012].

When the problem is represented using natural frequencies, an overall population size has to be selected. For instance, a population size of 1,000 could be used for the mammography problem in Section 6.2 as in Section 2.5.3. Though small populations (e.g., less than 100) are easier for participants to relate to [Brase, 2002] and to handle [Schapira et al., 2001], it is not always possible to use such a population size (e.g., if a population smaller than 1,000 is used for the mammography problem, some of the quantities would be fractions). Also, confidence in the presented information may be lost [Schapira et al., 2001] due to human intuition and the 'Law of Large Numbers' that larger samples are more reliable [Bernoulli, 1713; Sedlmeier and Gigerenzer, 1997]. However, when a large population size is used (e.g., 1,000), related risk could be perceived smaller than its actual magnitude [Schapira et al., 2001] (details in Appendix C.2.8). So the population size should be carefully selected and a different size for different problems should be used in a study to balance out the above effects.

Different User Abilities

A level of numeracy ability is required for the understanding and manipulation of natural frequencies [Chapman and Liu, 2009] and statistical information [Brown et al., 2011]. Spatial abilities are required to efficiently retrieve information from visuals and handle spatial properties [Kozhevnikov et al., 2002, 2005]. So a visualization may not facilitate reasoning for some [Gaissmaier et al., 2012] as individuals have different abilities [Galesic and Garcia-Retamero, 2010] (also Appendix C.2.9). In fact, Kellen's [2012] findings indicate that while Euler diagrams help people with high spatial abilities in contrast to text alone, those with low spatial abilities perform worse with an Euler diagram than with text alone. Thus studies should evaluate different visualization types as well as text alone and identify which representations, be they textual or graphical, are effective for different users types and levels of abilities.

Different representations could then be used to specifically aid different types of users [Spiegelhalter et al., 2011], but this is not practical in the real-world [Hawley et al., 2008]. The ideal alternative would be to have a representation that meets the needs of different types of users, but as mentioned earlier, this might only be possible for visual users. For instance, a study [Hawley et al., 2008] indicated that people with high numeracy benefit from any visualization and those with low numeracy benefit from visualizations with discrete objects such as frequency grids, so frequency grids could be used, despite that the success rate of those with low numeracy is lower than those with high numeracy. Similarly, Kellen et al. [2007] noted that people with high spatial abilities can handle any visualization but perform better with visualizations that have spatial rather than discrete properties, while those with low spatial abilities can only handle visualizations with discrete objects. So they hypothesized (but not evaluated) that frequency grids could be helpful for users with different spatial abilities. Garcia-Retamero and Hoffrage's [2013] findings indicate that a hybrid diagram made up of rectilinear closed curves filled up with square, touching glyphs that form a grid help doctors and patients be more accurate, even when the numeracy abilities of the participants are taken into account.

6.2.3 Our Objectives

Previous studies had limitations and their findings were inconsistent, so an effective representation of Bayesian problems, be it textual or graphical, is still unknown (Section 6.2.2). We assessed the effect of six visualizations (i.e., area-proportional Euler diagrams, Figure 6.2A-B; glyph representations, Figure 6.2C; Euler diagrams with glyphs, Figure 6.2D-F) and text alone on such problems to identify:

1. the most effective representation in general, be it professionals or laymen;
2. whether hybrid visualizations are helpful;
3. whether there is a link between the representation and different abilities (or user types).

To achieve these objectives, we conducted a study over three classic Bayesian problems on a large, diverse group of participants through crowdsourcing—the first crowdsourced study in the area.

Two novel textual problem formats, specially designed to be used with visualizations, were also proposed and evaluated. This makes it the first study to evaluate various types of visual and textual representations using the "same experimental methodology" [Brase et al., 2006, p. 974] and for different classic Bayesian problem. Other previous studies evaluated similar area-proportional Euler diagrams (Figure 6.2A—[Brase, 2009; Sloman et al., 2003]; Figure 6.2B—[Kellen, 2012]) and frequency grids (Figure 6.2C—e.g., [Cosmides and Tooby, 1996; Sedlmeier and Gigerenzer, 2001]), but none assessed our Euler diagrams with glyphs (i.e., Figure 6.2D-F). It is also the only known study in this area that

considered the participants' spatial as well as numeracy abilities (Kellen [2012] only spatial abilities; Garcia-Retamero and Hoffrage [2013] only numeracy abilities), which could affect performance. The rationale of the diagram designs is clearly explained (Section 6.4.2). The diagrams were accurately generated with respect to the quantitative data in the problem using our method *eulerGlyphs*, which we specifically devised to draw the evaluated visualization types for any Bayesian problems (Section 6.3).
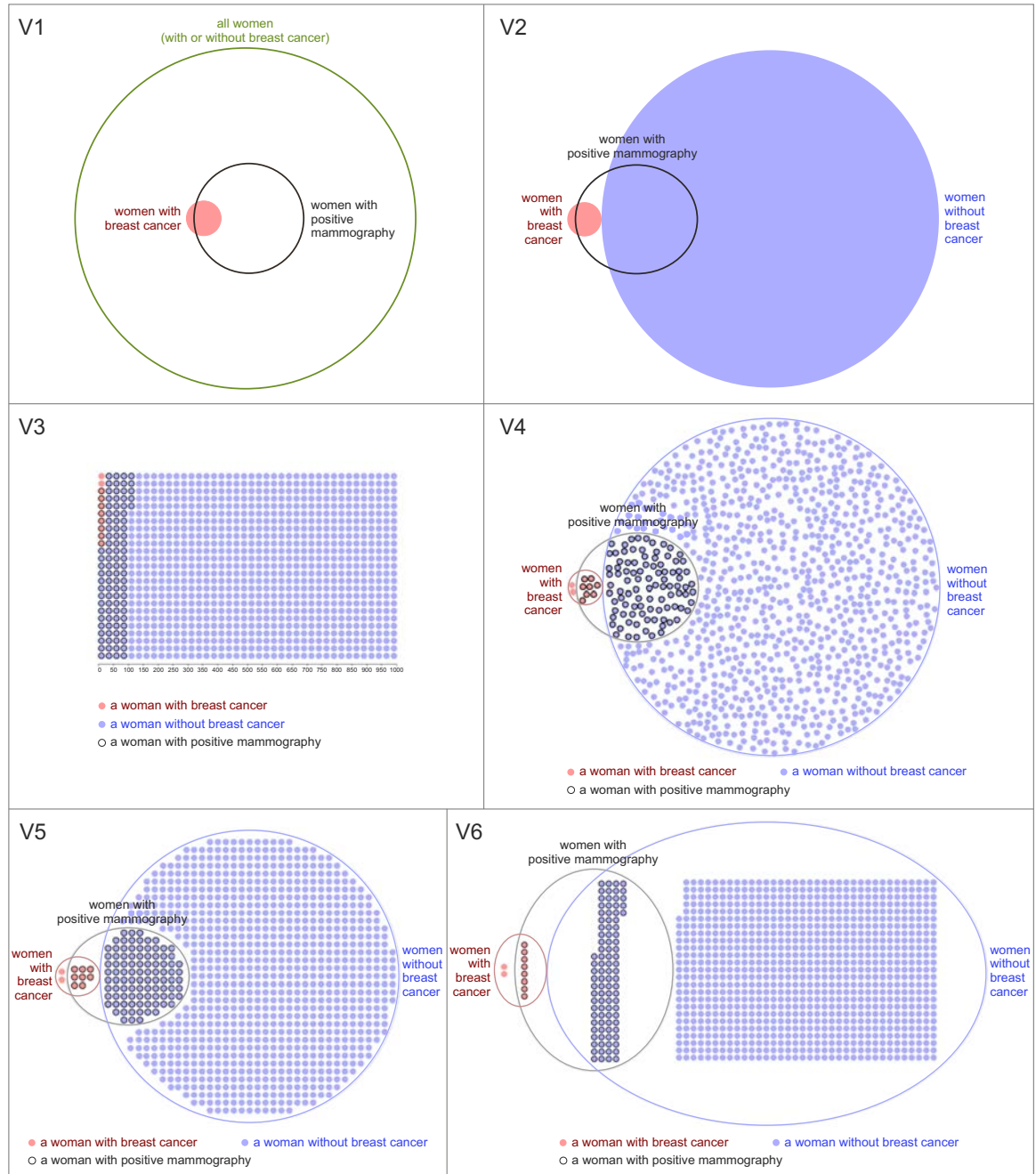
Figure 6.2: *The six visualization types evaluated in our study.* Here the visualizations (generated by *eulerGlyphs*) show the classic mammography problem [Eddy, 1982]. (V1) An area-proportional Euler diagram with the population represented as one set (the green curve). (V2) An area-proportional Euler diagram with the population represented as two sets (the red and blue curves). (V3) A frequency grid. (V4) An area-proportional Euler diagram with randomly positioned glyphs. (V5) A not area-proportional Euler diagram with uniformly positioned glyphs. (V6) A not area-proportional Euler diagram with frequency grid glyphs.

6.3 The Drawing Method – euler*Glyphs*

Given the base rate, the hit rate and the false alarm rate (as percentages) and the population size of a Bayesian problem, euler*Glyphs* generates the different types of visualizations in Figure 6.2. All the visualizations are consistent with one another. For instance, the shape and the size of the glyphs in V3-V6 and the distance between the glyphs in V3, V5, V6 is the same for the relevant visualizations. The labels and the colour coding scheme (both the solid and outline colours) for the sets are also consistent. Features, such as the labels of the curves and the shape of the glyphs, are yet customizable.

The Bayesian problems handled by euler*Glyphs* are comprised of two mutually exclusive and exhaustive hypotheses and the data obtained from one single observation [Gigerenzer and Hoffrage, 1995]. For instance, in the classic mammography problem [Eddy, 1982] (discussed in Section 6.2), the two hypotheses are *breast cancer* and *no breast cancer*, while the observation is *the positive mammography*. The task is to estimate the likelihood that one of the two hypotheses is true based on the data obtained from the observation [Gigerenzer and Hoffrage, 1995]. Visualizations such as euler*Glyphs*'s V1-V6 (e.g., Figure 6.2) could facilitate reasoning by illustrating the set relations and cardinalities.

In this section, the sets characterizing the Bayesian problem will be referred to as follows:

- S, the entire population considered in the problem (i.e., the sample space);
- H and $\overline{H}$, are the two mutually exclusive and exhaustive hypotheses, so $H \cup \overline{H} = S$;
- D, the data obtained from the single observation.

As illustrated in Figure 6.2, S is only depicted in V1 (in green), while $\overline{H}$ is only depicted in V2-V6 (in blue). H and D are depicted in all visualization types V1-V6 (in red and black respectively). Before drawing the actual visualizations, euler*Glyphs* must first compute the cardinality of the required sets and their intersections. We explain how euler*Glyphs* computes these values (Section 6.3.1) and draws visualization types V1-V6 to depict this quantitative data (Section 6.3.2).

This section explains *how* the visualizations are drawn. The rationale of the visualization designs in relation to Bayesian reasoning, that is *why*, is discussed later in Section 6.4.2.

The software euler*Glyphs* is available as freeware at `http://www.eulerdiagrams.org/eulerGlyphs`.

6.3.1 Computing the Cardinality of the Sets and Their Intersections

Percentages for the base rate, the hit rate and the false alarm rate are provided as inputs together with the population size of the problem. Based on these inputs, euler*Glyphs* computes the cardinality of the required sets and their intersections as follows:

- $|S|$ = *population size of the problem*;
- $|H|$ = *base rate* $\times$ $|S|$;
- $|\overline{H}|$ = $|S|$ - $|H|$;
- $|D \cap H|$ = *hit rate* $\times$ $|H|$;
- $|D \cap \overline{H}|$ = *false alarm rate* $\times$ $|\overline{H}|$.

Based on this quantitative data, euler*Glyphs* draws visualization types V1-V6 as explained in the next section.

6.3.2 Drawing the Visualizations

V1: Area-Proportional Euler Diagram with a 1-Set Population

V1 is the only visualization representing the entire population of the Bayesian problem as one set, *S*. It is an Euler diagram which depicts sets *S*, *H* and *D* as circles. The area of each circle is proportional to the cardinality of the depicted set, while the area of each zone is proportional to the cardinality of the set intersection it represents. So V1 is an accurate area-proportional Euler diagram with respect to the quantitative data in the Bayesian problem.

An accurate area-proportional Venn diagram with two circles can be drawn with respect to any set of zone areas using a numerical method (Section 3.7.5). So euler*Glyphs* uses Chow and Ruskey's [2004] bisection method (explained in Section 3.7.6) to draw an accurate area-proportional 2-Venn diagram for *H* and *D* and their intersection. This diagram is then placed inside a circle representing *S* whose area is proportional to the population size of the problem.

By default, the 2-Venn diagram for *H* and *D* is placed at the centre of the curve for *S*, but options are available for the diagram to be moved to the left, right, top or bottom of the interior of curve *S*. The curve for *S* is by default a circle, but an option is available to represent this as a square or rectangle.

V2: *Area-Proportional Euler Diagram with a 2-Set Population*

In contrast to V1, V2 (and V3-V6) represents the entire population of the Bayesian problem as two sets, *H* and *H̄*. Together with *H* and *H̄*, it depicts *D*. It is an Euler diagram drawn using two circles (for *H* and *H̄*) and a ellipse (for *D*). The area of each circle is proportional to the cardinality of the depicted set. The area of the overlap between each circle and the ellipse is also proportional to the intersection cardinality of the relevant sets. In fact, the area of all the zones, except for one, are proportional to the quantities in the Bayesian problem. The only zone whose area does not represent any data in the problem and is thus empty is the disconnected zone in the ellipse for *D* that is comprised of two regions, of which none are in the circle for *H* or *H̄*. This zone and its area are irrelevant to the problem and should be ignored.

Our method euler*Glyphs* draws V2 as follows:

1. An accurate area-proportional Venn diagram with two circles for *H* and *D* and another for *D* and *H̄* is drawn using Chow and Ruskey's [2004] bisection method.

2. The two diagrams are centre aligned vertically and moved close to one another along the *x*-axis until the circles for *H* and *H̄* touch at a single point or the two circles for *D* are merged into one, whichever occurs first. The two circles for *D* can now be replaced with an ellipse that depicts the cardinality of the intersections between *H* and *D* and *D* and *H̄*.
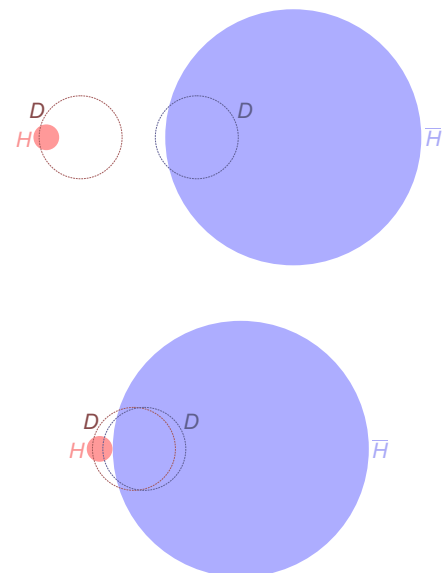
3. The centre of the ellipse for *D* is set to the midpoint between the centre of the two circles for *D*. The semi-horizontal axis of the ellipse is set to the distance between the centre of the ellipse and the leftmost intersection point of the circumference of the leftmost circle for *D* with the circle's horizontal diameter, so that the width of the ellipse covers the width of both circles for *D*. An appropriate semi-vertical axis for the ellipse is one that ensures that the area of the zones in curves *H* and $\overline{H}$ is proportional to the quantitative data of the problem. This is found using the bisection method in the interval [*l*, *u*], where

l is half the length of either the chord between the intersection points of circles *H* and *D* or the chord between the intersection points of circles *D* and $\overline{H}$ depending on which one is the longest,

and

u is the semi-horizontal axis of the ellipse.

Once the require semi-vertical axis is found, *D* is depicted as one curve and the area of all the zones relevant to the problem is proportional to the data of the problem (here the mammography problem [Eddy, 1982]) with a maximum relative error percentage of less than 1%.
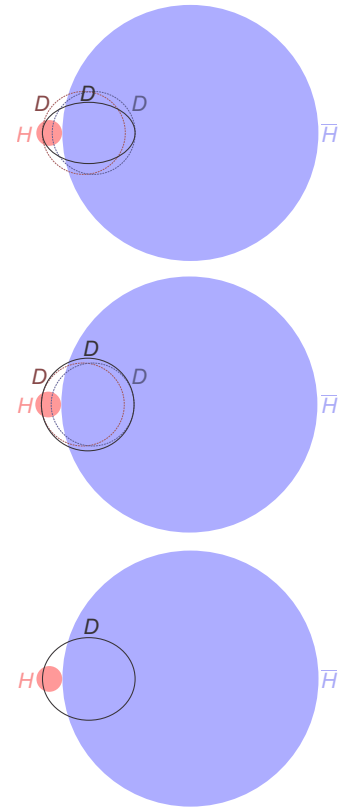
V3: *Frequency Grid*

V3 is a frequency grid depicting each member of the population of the Bayesian problem as a glyph. Similar to V2, sets *H, $\overline{H}$* and *D* are depicted. Every glyph is a member of one of the sets *H* or $\overline{H}$ and possibly *D*. Their solid and outline colours indicate their set memberships, but their shape and size are consistent within the visualization and among visualization types. A legend indicates the solid and outline colour assigned to every set, which colours are consistent with those of other visualization types.

The glyphs are can be laid out horizontally (row by row) or vertically (column by column) and in any direction (left to right or right to left; top to bottom or bottom to top; zig-zag or not). Example, in Figure 6.2, the glyphs of V3 are laid out vertically, left to right, top to bottom and in a zigzag manner so the second column is filled in bottom to top, the third top to bottom and so on. A ruler is provided to indicate the direction of the glyph placement and to aid estimation of the number of glyphs.

The number of glyphs per row and column is calculated, unless the user indicates a specific aspect ratio. If possible a square grid is used, else a grid with similarly sized rows and columns (e.g., multiples of fives or tens) is used (e.g., 20 rows, 50 columns for population 1000). First *n*, the square root of the population size, is calculated. If *n* is a whole number, a $n \times n$ square grid is used. Else, the number of glyphs per row is set to $5 \times 10^{s-3}$ where *s* is the number of significant figures of the population size.

By default, the glyphs are circles, but are customizable to squares or person figures, in which case, the shape of the glyphs of the other visualization types (i.e., V4-V6) is also changed. The size of the glyphs is also consistent among visualization types. The size is obtained by dividing the area of the largest inscribed rectangle in the smallest region of V2 by the population size, and then scaling the glyph so that it fits in this area together with the required spacing between the glyphs.

V4: *Area-Proportional Euler Diagram + Randomly Positioned Glyphs*

V4 is the same as V2 with the only difference that same sized glyphs are placed in random locations inside only the zones relevant to the problem, so no glyphs are added to the disconnected zone in the curve D but in none of curves H or $\overline{H}$. The number of glyphs in the zones is equal to the cardinality of the set intersection depicted by the zone. The glyphs in V4 are the same as those in V3, V5 and V6 in shape and size to ensure consistency and integrity. The size of the glyphs is automatically computed as discussed in the section for V3. Similar to V3, a legend for the sets is displayed below the diagram.

The glyphs are placed using an iterative random placement algorithm without packing. The largest rectangle enclosing the zone is first found and then a random point in the rectangle is selected. If the point is inside the zone and a glyph can be placed at that point such that it is inside the zone and it does not collide with other glyphs already in the zone, then a glyph is added. If not, another random point is selected. This iterative process continues until the required number of glyphs are added accordingly.

V5: *Not Area-Proportional Euler Diagram + Uniformly Positioned Glyphs*

V5 is the same as V4 except that the glyphs are placed in the form of a grid and the Euler diagram could be not area-proportional as the smallest curves that enclose the glyphs of the relevant zones are drawn.

The size of the glyphs and the spacing between the glyphs is consistency with other visualization types. So the area occupied by each glyph, including the required spacing, is known and will be referred to as *i*. This means that a zone *z* with *n* glyphs requires an area of *requiredArea(z) = i × n*. Knowing this, curves whose size is just right to only fit the relevant glyphs can be computed as follows:

1. The largest inscribed rectangle in the region of each zone *z* of V2 (the area-proportional Euler diagram based on which also the glyph size is computed) is obtained and its area is referred to as *currentRectArea(z)*. Each curve *c* in V2, with *Z* as the set of zones located in *c,* is then scaled by

$$\sum_{z \in Z} requiredArea(z) \bigg/ \sum_{z \in Z} currentRectArea(z)$$

2. After a curve is scaled, the other curves in the diagram are displaced accordingly to ensure that the same set intersections as in V2 are depicted.
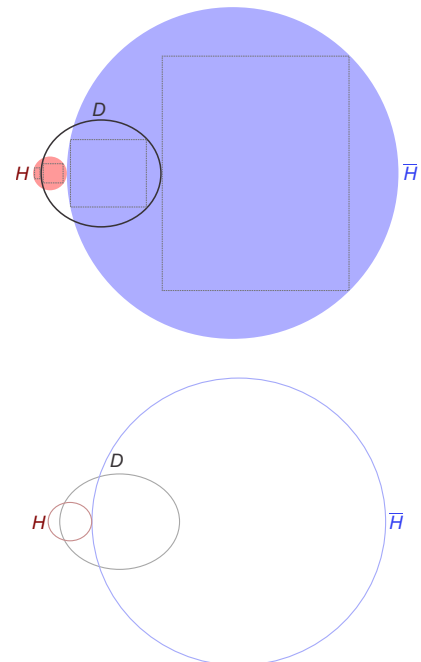
3. The largest inscribed rectangle in the region of each zone of the modified diagram with scaled curves is then obtained.
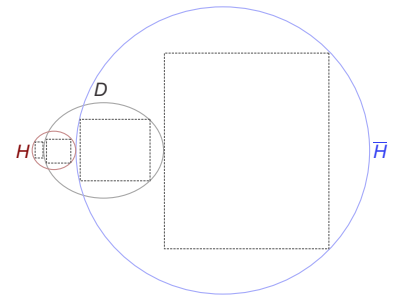
4. Glyphs are placed within the zones by first filling up the inscribed rectangles.
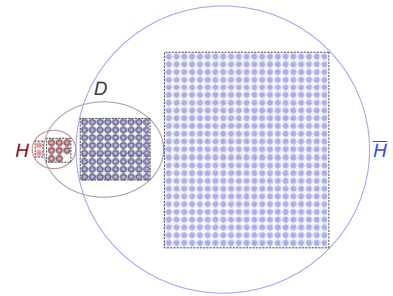
5. Other glyphs are then placed around the sides of the inscribed rectangle ensuring that all the glyphs are inside their respective zone and aligned in the form of a grid.
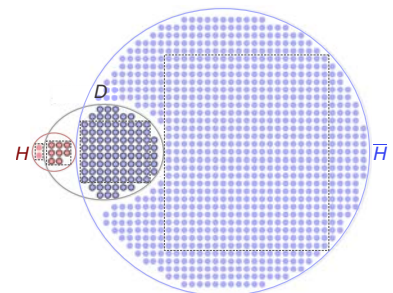
6. Once all the glyphs are added, the inscribed rectangles are removed.
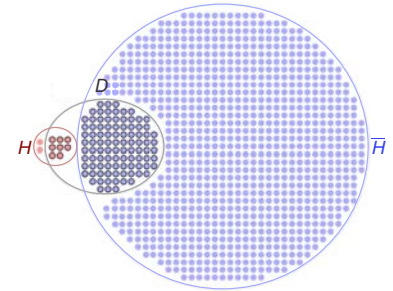
    Here the diagram depicts the quantitative data of the mammography problem [Eddy, 1982].

V6: *Not Area-Proportional Euler Diagram + Frequency Grid Glyphs*

V6 takes the frequency grid in V3, moves the groups of glyphs apart and adds not area-proportional curves to depict the sets, their intersections and glyph set memberships. This is drawn as follows:

1. Frequency grid V3 is generated and the four different groups of glyphs are moved apart and centre aligned vertically.
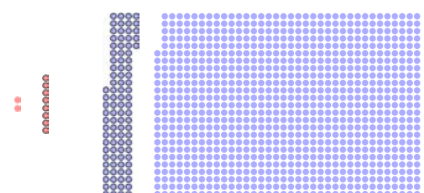
2. The two groups of glyphs that are members of H and the two others that are members of $\overline{H}$ are moved further apart from one another and bounded by an ellipse

that circumscribes the minimal bounding box of the two groups.

3. A similar ellipse is drawn around the two groups of glyphs that are members of D.

This ellipse for D might intersect one or both groups of glyphs that are in H or $\overline{H}$ but not in D.

In such cases, the relevant group of glyphs is moved away to the left (if its glyphs are members of H) or right (if its glyphs are members of $\overline{H}$) until it is out of the curve for D.

If it is moved n units, both semi-axes of the ellipse depicting the set its glyphs are members of are enlarged by n. The x-coordinate of the centre of this ellipse is decreased or increased (depending on the direction of the initial move) by $n/2$.

Here the diagram depicts the quantitative data of the mammography problem [Eddy, 1982].
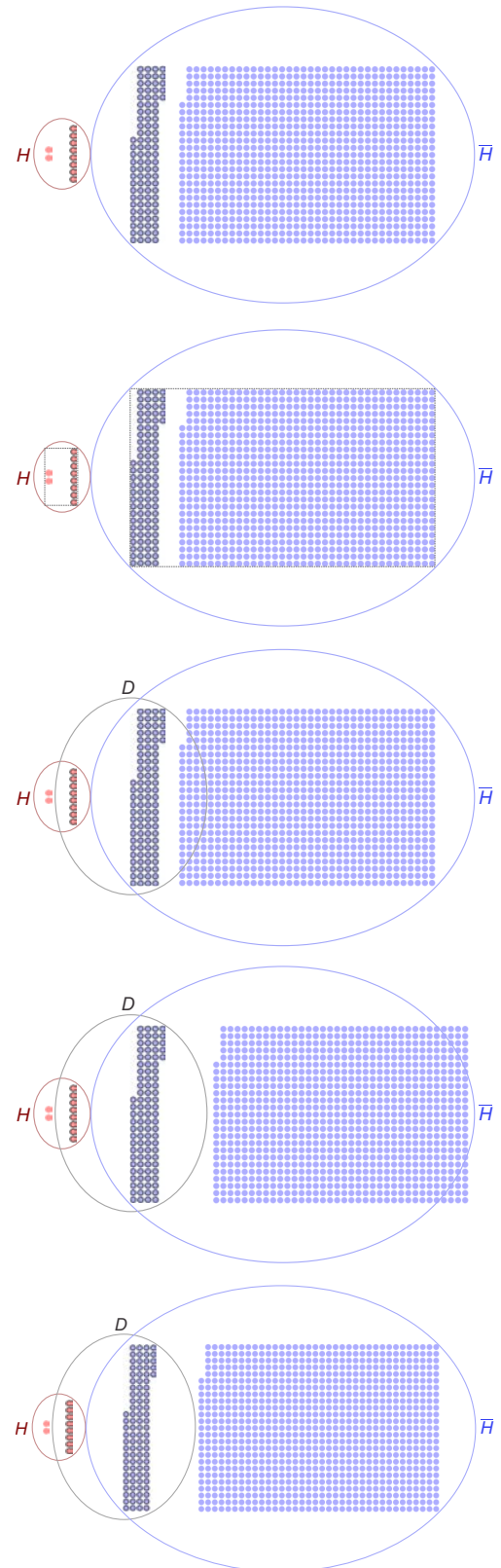
6.4 Study Design Rationale

After devising *eulerGlyphs*, we conducted our study to assess the effectiveness of the visualizations for Bayesian reasoning. In this section, we motivate our study design, including the use of crowdsourcing, our different visualization designs, our choices of Bayesian problems and our performance measures.

6.4.1 Crowdsourcing

Most studies on Bayesian reasoning were carried out on populations with a specific background, often students. This poses problems in term of generalizability and ecological validity (Section 6.2.2 subsection 'Participants') as in the real-world both laymen and professionals are faced with probabilistic, uncertain data based on which they have to make important timely decisions, possibly with limited cognitive abilities that are known to decline with age [Hale et al., 2007; Salthouse, 1996].

For these reasons, we considered crowdsourcing and we used Amazon Mechanical Turk (MTurk) as a technology to automatically outsource simple tasks to a network of Internet users [Chen et al., 2011; Paolacci et al., 2010]. The tasks posted by *requesters* are called *HITs* (Human Intelligence Tasks) and are completed by anonymous *workers* who get a monetary reward, if they complete the task appropriately. Crowdsourcing platforms have not been initially designed for conducting experiments, but their use in research is popular [Chen et al., 2011; Horton et al., 2011], including in information visualization [Heer and Bostock, 2010]. The workers' demographics is well-understood [Ross et al., 2010] and a methodology is being developed for designing effective experiments and addressing concerns such as scientific control [Heer and Bostock, 2010; Horton et al., 2011; Paolacci et al., 2010].

Although crowdsourced experiments are subject to many of the same problems as laboratory experiments, they capture interesting aspects of real-world problem solving. First, they capture a large, diverse population with different backgrounds, education, occupations, age groups and gender [Ross et al., 2010]. Secondly, workers typically try to complete as many HITs as possible (often for personal satisfaction), while a rating system provides them with incentives for being accurate [Horton and Chilton, 2010; Paolacci et al., 2010]. Since workers typically complete several HITs in sequence, they cannot focus on a single task like participants in a laboratory. We believe this better captures situations when timely decisions have to be made accurately. Also, due to the informal setup, participants might be less subject to experimental biases like demand characteristics [Nichols and Maner, 2008].

Finally, setting up an experiment on a crowdsourcing platform can be initially costly, but the time and effort for running subjects is much lower than in laboratory experiments. In contrast to previous studies, experiments should be conducted on diverse, large populations to ensure generalizable findings, detectable small effects and replicable results (Section 6.2.2). Crowdsourcing gives access to more statistical power and makes it easier to test multiple, diverse hypotheses as those involving equivalence.

6.4.2 Visualization Designs

Previous studies investigated whether visualizations can aid Bayesian reasoning, but only one or a few visualizations were evaluated in every individual study (Section 6.2.2 subsection 'Diagram Designs'). In addition, the choices made in terms of visualization design were rarely discussed and often inconsistent with respect to the quantitative data they are meant to represent. Overall, this impedes understanding and comparison of findings of different studies.

To address this, we assessed a set of visualization designs (V1–V6; Figure 6.2 illustrates the diagrams for the classic mammography problem [Eddy, 1982]; the diagrams for the other two classic Bayesian problems considered in our study are in Appendix B.1) including Euler diagrams, glyph-based representations and combinations of both. Consistent with the tradition of research in information visualization and human-computer interaction, our goal was to start paving the design space based on clear design rationales, while keeping unimportant design details as consistent as possible in order to better tease out the effects of important design features. The latter is important as slight changes in the design could affect the subject's perception of the data semantics [Ziemkiewicz and Kosara, 2010b].

Visual design features that are known to hinder understanding such as non-smooth and irregular curves for Euler diagrams (discussed in Section 3.5 for not area-proportional Euler diagrams and in Section 3.7.3 for area-proportional ones) were avoided whenever possible. All the colours used in our visualizations were obtained from one of ColorBrewer's [Harrower and Brewer, 2003] colour-blind friendly scheme. We also used Adobe Illustrator to preview how people with different types of colour-blindness would view our diagrams to ensure that the selected colours are, in such cases, distinguishable as much as possible. Prominent data sets that appear in most of the visualizations were depicted using colours (namely, red and blue) that are highly distinguishable for not colour-blind and colour-blind of any type. All the visualizations were accurately generated by euler*Glyphs* (Section 6.3).

We strived to make our visualizations compact, clear and consistent with the numeric data of the problem ("To minimize the cost of visual searches, make visualization displays as compact as possible, compatible with visual clarity" [Ware, 2012, p. 141 G5.1]) be it represented as area (area "should be directly proportional to the numerical quantities represented" [Tufte, 1983, p. 56]) or glyphs (the number of glyphs is equal to the quantity). All the elements of our visualizations were clearly labelled to ensure graphical integrity ("Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity" [Tufte, 1983, p. 56]).

V1: *Area-Proportional Euler Diagram with a 1-Set Population*

V1 (Figure 6.2A) is an accurate area-proportional Euler diagram with respect to the quantities of the Bayesian problem. Similar to Brase's [2009] (Figure 2.9A) and Sloman et al.'s [2003] (Figure 6.1A) Euler diagrams, the entire population of the problem is represented as one set. We depict this set as a green outlined circle (Figure 6.2A). The area of the red circle is proportional to the base rate. The interior of the black outlined circle is split up into two regions: one (in red) whose area is proportional to the hit rate; another (in white) whose area is proportional to the false alarm rate. The answer to a Bayesian problem is the area of the hit rate region as a fraction of the area of the black circle.

V1 is the only diagram depicting the entire population of the Bayesian problem as one circle and thus, its design unavoidably differs from the other visualizations (V2–V6). The other diagrams depict the entire population as two circles, one shaded in red and another in blue. The entire population in V1 could be represented as two sets as in Figure 6.3, but an early pilot study revealed that this could be misleading as subjects assumed that the red circle was part of the blue region. So instead of seeing the blue as a circle with a hole, subjects thought the blue was an entire circle with the red circle as a subset.

Thus, we opted to show the entire population as a circle as in Figure 6.1A. We wanted a style and colour that is distinguishable from the data sets depicted in red and blue, even for colour-blind subjects, and so we chose a green outlined circle.
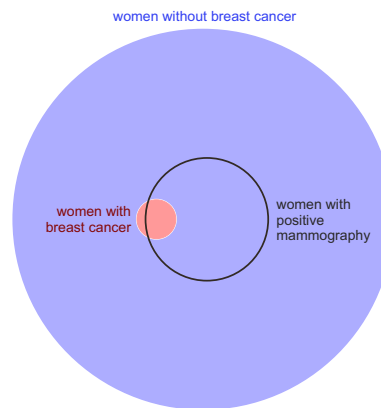
Figure 6.3: *An alternative design of V1 that is more consistent with the other visualization (V2–V6).*

V2: *Area-Proportional Euler Diagram with a 2-Set Population*

V2 (Figure 6.2B) is an accurate area-proportional Euler diagram, except for a disconnected empty zone (the white space in the black ellipse). V2 is different from V1 as it represents the entire population as two sets (the red and blue circles in Figure 6.2B). A similar design was proposed [Kellen et al., 2007], but only evaluated (Figure 3.37) [Kellen, 2012] after our study [Micallef et al., 2012] was published.

To reinforce the complementarity of the two population sets, two disjoint yet touching circles with similar styles and contrasting colours that help relate their content are used. For instance, for the mammography problem, the red circle represents the group of 'women with breast cancer' and the blue circle represents the group of 'women without breast cancer'. Also, red and blue are: unique hues [Ware, 2008, p. 70], thus distinct and contrasting [Mullen, 1985]; primary colours, so understood across cultures [Berlin and Kay, 1969]; visually processed in parallel on distinct channels [Hurvich, 1981]. For the Bayesian cab problem [Tversky and Kahneman, 1982] which we also tested, we replaced green and blue cabs in the original problem with red and blue cabs in our experiment to ensure consistency with the design and colour scheme (that is also colour-blind friendly) of our visualizations.

An incorrect answer that is much bigger than expected is often provided for the mammography problem because the base rate, which is low, is ignored. The area of the red circle is proportional to the base rate and so, seeing it next to the blue circle could reduce the likelihood that it is ignored or overestimated. Also, comparing the area of two circles is easier than that of two irregular shapes [Chow and Ruskey, 2004] and have thus been used by Playfair [1801] (e.g., for populations—Figure 2.2A), as map symbols [Montello, 2002], and in a geological survey to show the "relative amounts of Earth's water in comparison to the size of the Earth" [Perlman, 2013]. Thus, by depicting the population as two sets, we anticipated that compared to V1, V2 would clarify the nested-set relations of the problem and aid Bayesian reasoning. Since the third set depicts a different concept, a different shape (an ellipse) is used and only its outline is displayed, making it distinguishable from the two population sets [Ware, 2008, 2012]. As discussed in Section 3.5 and Section 3.7.3, circles (followed by ellipses) are often preferred for the curves of Euler diagrams due to their regularity and smoothness. Thus, their use in V2.

A drawback of V2 is that it has a zone that is not relevant to the Bayesian problem and is thus, empty. This zone is also disconnected as it is made up of the two white regions that are inside the black ellipse. So for the mammography problem, this zone corresponds to the set of 'women with a positive mammography', but that belong to none of the two population sets. To avoid having this empty zone,

the third set could be depicted as a self-intersecting curve (Figure 6.4A) or as two circles with the same label (Figure 6.4B). However, these diagrams are non-wellformed (Section 3.4.4) as Figure 6.4A has a self-intersecting (i.e., non-simple) curve and multiple points and Figure 6.4B has duplicate curve labels. To depict the third curve as a smooth curve and avoid the empty zone, the two population sets could be depicted as rectangles. However, such diagrams cannot be drawn accurately for all data. Example for the mammography problem (Figure 6.5A), the area of the red and blue rectangles and that of the black ellipse are proportional to the required quantities, but the area of the regions inside the black ellipse are not, as it is not possible to move the ellipse further to the left. If alternatively the third set is another rectangle (Figure 6.5B), the diagram would be non-wellformed and unusable [Rodgers et al., 2012b] due to concurrent curves (Section 3.5). So the diagrams in Figure 6.4 and Figure 6.5 should be avoided.

The design in Figure 6.2B seems more familiar and easier to understand than other alternatives. For instance, any woman that is in the set of 'women with a positive mammography' must also be in only one of the sets 'women with breast cancer' or 'women without breast cancer', so the empty zone is easily ignored. This is easier in V4-V6 as glyphs are added to all the regions except to the empty zone. V2 is non-wellformed due to the brushing point between the red and blue circles, but such points do not have negative effects on the diagram's comprehension [Rodgers et al., 2012b].
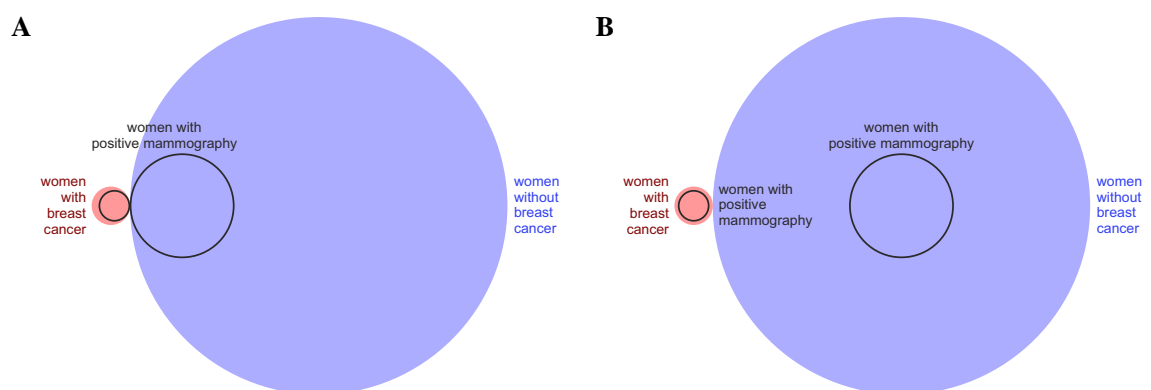
Figure 6.4: *Possible alternative designs for V2 with no empty zones.*

Figure 6.5: *Possible alternative designs for V2 that depict the two population sets as rectangles.*

V3: *Frequency Grid*

V3 (Figure 6.2C) maps equally-sized circular glyphs to different sets in the Bayesian problem, with as many glyphs as members of the population. The design is representative of typical designs in the literature on risk communication, including the horizontal ruler provided to aid cardinality estimation.

The same colour scheme (solid and outline colours) as in V2 is used to convey set membership of glyphs. This ensures consistency and integrity. Ware advises to use unique hues for small symbols [2012, p. 123 G4.11], so the colours used in V2 are also appropriate for the glyphs in V3 and V4-V6 which also have glyphs. A study suggests that since hue and orientation are preattentive features, the number of glyphs that only vary by hue and orientation can accurately and rapidly be estimated [Healey et al., 1996]. We are using distinct hues to represent set membership and so the number of glyphs belonging to the same set should be accurately and rapidly estimated.

Glyphs of different shapes have been used before, including simple geometrical shapes [Galesic et al., 2009] and icons such as anthropomorphic figures [Brase, 2009; Brown et al., 2011]. Studies report no significant improvement of icons that are visually similar to the objects they represent over simple shapes [Stone et al., 1997] and no effect on comprehension and recall [Gaissmaier et al., 2012]. In addition, icons are problem-dependent and can make diagrams cluttered [Galesic et al., 2009]. Hence, we opted for circles. Studies assessing Euler diagrams with glyphs for Bayesian reasoning have also used simple shapes like circles [Brase, 2009] or squares [Garcia-Retamero and Hoffrage, 2013].

Glyphs in a grid that are represented as simple geometric shapes (e.g., circles or squares) and that are very close to one another, can be visually processed in a discrete manner or as area [Ancker et al., 2006]. Thus, meeting the preferences and abilities of different readers (those with low spatial abilities prefer discrete objects; those with high spatial abilities prefer area [Kozhevnikov et al., 2002, 2005]). Also, studies indicate that humans are biased to area judgement and quantities should not by area but by representations that involve tasks that humans are better at, like position or length judgement along a common scale (Section 3.7.3), both of which are possible with such grids of equally-sized glyphs.

Different layouts have been proposed in the literature. Sometimes, glyphs are laid out horizontally [Brown et al., 2011], vertically [Price et al., 2007] or randomly [Brase, 2009]. A study [Price et al., 2007] suggests that horizontal grids are perceived faster. Brase [2009] argues that sequential and random placement are both effective, but others [Ancker et al., 2011] suggest that randomness increases subject's uncertainty as proportions in the diagram are harder to estimate, differences are less easily noticeable, and larger proportions are perceptually overestimated. We therefore used a sequential layout, but placed glyphs vertically using an ordering for the sets that matches the layout of V2.

For the grid dimensions, we used aspect ratios that are typical in the literature: a 25×40 grid (for problems with a population of 1000) and a 10×10 grid (for problems with a population of 100). Since glyphs are difficult to label in-place, a separate legend was provided.

V4: *Area-Proportional Euler Diagram + Randomly Positioned Glyphs*

V4 (Figure 6.2D) consists of the area-proportional Euler diagram V2 with glyphs randomly positioned in its regions. It is similar to Brase's [2009] diagram (Figure 3.39A), except that our area-proportional Euler diagram is accurate with respect to the quantities of the problem (excluding the disconnected, empty zone in the black ellipse) and the number of glyphs matches the values in the problem.

The size and shape of the glyphs for V4 (also V5, V6) is consistent with V3. A legend (for glyphs) and in-place labels (for Euler curves) are provided. The same contrasting colours and style as in V2 are used for the glyphs and the outline of the transparent Euler curves, so they are distinguishable and also easily visible in the Euler diagram ("To make symbols easy to find, make them distinct from their background and from other symbols …" [Ware, 2012, p. 149 G5.3]; "For small color-coded symbols,

ensure luminance contrast with the background as well as large chromatic differences with the background." [Ware, 2012, p. 123 G4.12]). Hence, regions with no glyphs look empty and irrelevant, and the coloured glyphs pop out [Ware, 2012, p. 155]. Also, a study indicates that searching for an empty convex shape among a number of convex shapes with a dot inside is slower than searching for a convex shape with a dot outside among a number of convex shapes with a dot inside [Treisman and Gormican, 1988] as the "search for the presence of a visual primitive is automatic and parallel, whereas search for the absence of the same feature is serial and requires focused attention" [Treisman, 1985, p. 156]. So the regions with glyphs are automatically processed, while those with no glyphs are noted only through focused attention. Thus, the empty unwanted zone is easily ignored.

V4 and the other hybrid visualizations (V5, V6) could be more effective than V3, as elements that are enclosed by a curve tend to be perceptually grouped together (the perceptual grouping principle of common regions, which is stronger than others [Koffka, 1935] such as the Gestalt proximity and similarity principles) [Palmer, 1992] and closed curves are processed preattentively, thus noted without focused attention [Enns, 1986; Treisman and Souther, 1985]. In fact, one of Ware's design guidelines advises to "Consider putting related information inside a closed contour" [Ware, 2012, p. 187 G6.5].

V5: *Not Area-Proportional Euler Diagram + Uniformly Positioned Glyphs*

V5 (Figure 6.2E) is similar to V4 but employs a regular grid layout with the same glyph spacing as in V3. The smallest curves that enclose the glyphs are drawn after the glyphs are positioned in the correct regions. Thus, the Euler diagram is more compact but not necessarily area-proportional.

V6: *Not Area-Proportional Euler Diagram + Frequency Grid Glyphs*

While V4 and V5 are 'Euler diagram oriented' as glyphs seem to be added to the Euler diagram in V2, V6 (Figure 6.2F) is 'frequency grid oriented' as Euler diagram curves seem to be added to the frequency grid in V3. Hence, no attempt is made to ensure that the Euler diagram (in V6) is area-proportional.

Brase's [2009] study involved a similar representation (Figure 2.10A), but instead of moving apart the different groups of glyphs in the frequency grid to ensure that the Euler diagram curves are smooth and regular (as in our V6), Brase just added a rectangular closed curves over the frequency grid and highlighted some glyphs in the grid to indicate other set memberships. Brase's representation could be more compact than our V6. However, his representation could be difficult to comprehend as the curves are not smooth [Benoy and Rodgers, 2007] and no labels or legend is provided to understand what the curve and the highlight glyphs represent. Also, for the quantitative data of some Bayesian problems, the diagram could be cluttered and undesirable, non-convex, irregular curves might have to be used.

6.4.3 Bayesian Problems

Various previous studies involved a unique Bayesian problem. As discussed earlier in Section 6.2.2 subsection 'Bayesian Problems and Their Representation', this could be problematic. For instance, individual differences among participants such as interests, background, education, skills and experience could affect their adherence to and understanding of the problem [Brown et al., 2011; Davis et al., 2001]. Studying more than one problem can help level out such effects and facilitate the generalization of the experiment findings.

Gigerenzer and Hoffrage's [1995] classic study on Bayesian reasoning (that assessed textual but no visual representations) was carried out over 15 problems, each involving a different scenario. Three of these problems were classic Bayesian problems that had been considered in various previous studies, thus ensuring that their results were comparable to those of previous studies.

Crowdsourcing experiments have to be short [Heer and Bostock, 2010; Paolacci et al., 2010] and thus, we opted for the three classic problems that were considered in Gigerenzer and Hoffrage's [1995] study. These include:

- *Mam*, the mammography problem [Eddy, 1982];
- *Cab*, the cab problem [Tversky and Kahneman, 1982];
- *Eco*, choosing a course in economics problem [Ajzen, 1977].

The text, which we used in our study (mainly for Experiment 1—Section 6.5), for these three Bayesian problems is provided in Table 6.1. The evaluated visualizations (V1-V6) for each of these Bayesian problems are in Appendix B.1 (those for *Mam* are also available in Figure 6.2).

We opted for these specific problems as they have been tested by various previous studies (e.g., [Garcia-Retamero and Hoffrage, 2013; Gigerenzer and Hoffrage, 1995; Hoffrage and Gigerenzer, 1998; Kellen, 2012; Mellers and McGraw, 1999; Sedlmeier and Gigerenzer, 2001]—*Mam*; [Gigerenzer and Hoffrage, 1995; Kellen, 2012; Sedlmeier and Gigerenzer, 2001]—*Cab*; [Gigerenzer and Hoffrage, 1995; Sedlmeier and Gigerenzer, 2001]—*Eco*) and their scenarios are diverse.

Differences in the quantitative data used in different studies for the same Bayesian problem could make results less comparable (see Section 6.2.2 subsection 'Quantitative Data of Bayesian Problems'). Thus, we decided to use the same base rate, hit rate and false alarm rate as those used in the original classic problem. However, these values should ideally vary in distance and size to counterbalance any possible distance and size effect [Moyer and Landauer, 1967] (discussed in Section 6.2.2 subsection 'Quantitative Data of Bayesian Problems'). Our chosen classic problems satisfy this desirable characteristic as their values (which are the same as those in Gigerenzer and Hoffrage's [1995] study and consistent with the original proposed problem) are sufficiently varied, as noted in Table 6.2.

Various studies (e.g., [Gigerenzer and Hoffrage, 1995]) indicate that representing the problem's quantitative data as natural frequencies rather than probabilities significantly improves performance (Section 2.5.3). Thus, we opted for natural frequencies. The size of the problem's population can affect the participants' performance (e.g., perceiving relative risk smaller than its true size [Schapira et al., 2001]; discussed in Section 6.2.2 subsection 'Quantitative Data of Bayesian Problems') and so, the tested problems should ideally have a different population size. Our chosen classic problems satisfy this desirable characteristic as their population size (which is the same as that in Gigerenzer and Hoffrage's [1995] study) is diverse (i.e., 1,000 for *Mam* and *Eco*; 100 for *Cab*—Table 6.2).

Results of different studies could be less comparable, if the text used in the studies for the same Bayesian problem is different (discussed in Section 6.2.2 subsection 'Bayesian Problems and Their Representation'). To avoid this, we aimed to use the same text as that used in previous studies. The original text of these Bayesian problems used probabilities rather than natural frequencies and so, they could not be used for our study. Gigerenzer and Hoffrage [1995] only provided the text for the mammography problem, which we used for *Mam*. The text for *Cab* was obtained from another study that considered the same problem with the same quantitative data [Mellers and McGraw, 1999]. However, for *Eco*, no text in natural frequencies was found (the original used probabilities) and so, we

wrote our own text using the original [Ajzen, 1977] and a natural frequency style similar to that used for the other problems. The text for our problems is provided in Table 6.1.

Table 6.1: *The text used for the three Bayesian problems in our study, mainly for Experiment 1.*

Mam	10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.
	Here is a new representative sample of women at age forty who got a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer?
Cab	A cab was involved in a hit-and-run accident at night. Two cab companies, the Red and the Blue, operate in the city. Of every 100 cabs in the city, 15 are Blue and 85 are Red. On the night of the accident, a witness identified the cab as Blue. The court tested the reliability of the witness under the similar visibility conditions with Blue and Red cabs. When the cabs were really Blue, the witness said they were Blue in 12 out of 15 tests. When the cabs were really Red, the witness said they were Blue in 17 out of 85 tests.
	What are the chances that the cab involved in the hit-and-run accident was Blue?
Eco	In a small liberal arts college students take, as an elective, a general interest course in economics or history. A recent analysis of enrollment figures showed that out of every 1,000 students, 700 students took the general interest course in history, while 300 students took the course in economics. For 210 of the students out of the 300 who took the economics course, the decision on pursuing the economics course was career oriented. For 350 of the students out of the 700 who took the history course, the decision on pursuing the history course was career oriented.
	Barbara T. was one of the students who took one of the two general interest courses. Her decision on pursuing the course was career oriented. What are the chances that she had taken the economics course?

(*Mam*) Mammography problem ([Eddy, 1982]—original problem; [Gigerenzer and Hoffrage, 1995]—source text). (*Cab*) Cab problem ([Tversky and Kahneman, 1982]—original problem; [Mellers and McGraw, 1999]—source text). (*Eco*) Choosing a course in economics problem ([Ajzen, 1977]—original problem; we wrote the text based on the original problem).

Table 6.2: *The quantitative data used for the Bayesian problems in our study.*

	Base Rate	Hit Rate	False Alarm Rate	Answer to Question		
Mam	10 out of 1,000	8 out of 10	95 out of 990	8 out of 103	*or*	7.8%
Cab	15 out of 100	12 out of 15	17 out of 85	12 out of 29	*or*	41.4%
Eco	300 out of 1,000	210 out of 300	350 out of 700	210 out of 560	*or*	37.5%

These values are consistent with the original problems and are the same as those used in Gigerenzer and Hoffrage's [1995] study. (*Mam*) Mammography problem [Eddy, 1982]. (*Cab*) Cab problem [Tversky and Kahneman, 1982]. (*Eco*) Choosing a course in economics problem [Ajzen, 1977].

6.4.4 Measures of Performance

Although most previous studies focus on maximizing and reporting the proportion of correct answers, this dichotomous approach has some limits. First, a percentage of exact answers (say, 75%) says nothing about how far off the remaining 25% are. Second, it is rare that the outcome of a decision depends on a probability estimation being perfectly exact or not (e.g., making an estimation of 0.001 instead of the correct 0.0012), whereas an estimate of 0.4 versus 0.001 will often produce radically

different outcomes. Finally, helping people compute an exact answer might be useful in some situations (e.g., teaching probabilities), but is of limited relevance to many real-life situations where one has to make quick decisions and rarely has attentional resources and time to use pen and paper or a calculator.

We therefore chose to focus on accuracy, that is, how far subjects are from the actual answers. Since in the natural frequency format answers are provided as a nominator and a denominator (i.e., v_1 out of v_2), we start by computing the probability $p = v_1/v_2$. We believe this is acceptable since ultimately, the answer is a probability (e.g., the chances of having a cancer). We then compute a *bias*, which gives the error together with the direction of the error. Although one could use $p-p_e$, with p_e being the exact answer, subtracting probabilities can lead to paradoxes. For example, if $p_e = 0.01$, then $p = 0.0000001$ would be a more correct answer than $p = 0.02$. So we use $log_{10}(p/p_e)$ instead. The log makes it easy to report and compare large estimation errors. Thus, if $p_e = 0.01$ and $p = 0.001$, then the bias will be -1 and if $p = 0.1$ the bias will be 1.

We derive the *error* from the bias by computing its absolute value. This assumes that a negative bias is as serious as a positive bias, which is a reasonable assumption for problem-independent studies such as our user study. Alternatively, this measure can be adapted to situations where an overestimation is more costly than an underestimation or vice-versa, by pre-multiplying positive or negative biases with a constant. To be able to compare our data with previous experiments, we still report the occurrence of answers for which *error* = 0.

In addition to bias and error, we decided to measure the subjects confidence in their answers. This is important because if a visualization makes people very accurate but not confident at all, then it is of limited use. In contrast, a visualization that makes people very confident but plain wrong is harmful, more so than a visualization that makes people less accurate but not overconfident.

Finally, we also decided to measure the time spent reading problems and providing an answer. If workers devote very little time compared to, example laboratory experiment participants, then it could mean that they are not carrying out the task seriously. On the other hand, if they devote too much time, then maybe crowdsourcing does not capture quick real-world decision-making situations at all.

6.4.5 Measures of Abilities

As discussed in Section 6.2.2 subsection 'Different User Abilities', users have different numeracy and spatial abilities. Yet both are required to understand statistical information and visual representations. Frequency grids have been thought to effectively meet the needs of people with different levels of numeracy [Hawley et al., 2008] and spatial abilities [Kellen et al., 2007]. However, they might not be the best representation for all. For instance, those with high spatial abilities would benefit most from a visualization with spatial properties [Kozhevnikov et al., 2002, 2005], like Euler diagrams. Others, like those with low spatial abilities, find text alone more helpful than an Euler diagram [Kellen, 2012]. So we must identify which representation is best for different abilities and user types.

Thus, we decided to measure the numeracy and spatial abilities of our participants. Numeracy was measured using Brown et al.'s [2011] 6-question objective numeracy test due to the similarities of the subjects' demographics. To this, we added part 2 of the Subjective Numeracy Scale (SNS) [Fagerlin et al., 2007b]. Spatial abilities were measured using part 1 of the Paper Folding Test (VZ-2) [Ekstrom et al., 1976]. These are all paper-based tests that we faithfully reimplemented in HTML and JavaScript, including the 3-minute limit for VZ-2.

6.5 Experiment 1: Comparison of Visualizations

The purpose of this first experiment was to test the six visualizations (Section 6.4.2) and compare them with text alone. We hypothesized that our visualizations will help subjects solve Bayesian problems. Most of our visualizations were based on a less common Euler diagram representation for Bayesian problems because we anticipated that, by representing the population as two disjoint sets, the reader will be less likely to disregard the base rate. Following previous theories [Kozhevnikov et al., 2002, 2005], we also hypothesized that subjects with low spatial abilities would benefit from visualizations with discrete and countable objects such as glyphs (as in V3-V6), while those with high spatial abilities will benefit mostly from visualizations with spatial properties like Euler diagrams (as in V1-V2 and V4-V6).

6.5.1 Design

As experimental conditions, we had three Bayesian problems and seven visualization types namely text alone (V0) and the visualizations in Section 6.4.2 (V1-V6). Thus, our independent variables were:

- Bayesian problem: PROBLEM ∈ {*Mam*, *Cab*, *Eco*};
- Visualization type: VIS ∈ {V0, V1, V2, V3, V4, V5, V6}.

Our dependent variables were:

- BIAS, the difference between the subject's answer and the exact answer, computed as a log ratio;
- ERROR, the absolute value of BIAS;
- EX ∈ {0, 1}, whether the answer is exact or not;
- TIME, the time taken to solve the problem;
- CONF ∈ [1 .. 5], the subject's confidence in his or her answer.

Our covariates were:

- NUM ∈ [0 .. 30], the subject's score in the numeracy test;
- SPAT ∈ [0 .. 10], the subject's score in the paper folding task.

We used a mixed-design approach where each participant was presented the three problems, each accompanied by the same visualization type. The use of a between-subjects design for the visualization factor is consistent with previous studies (e.g., [Brase, 2009; Cosmides and Tooby, 1996]) and prevents asymmetric skill transfer effects [Poulton and Freeman, 1966]. To counterbalance any possible learning effect across problems, all the six possible orderings of the three problems were used. We had 24 participants per visualization and thus, each of the six problem orderings had one of the seven visualization types and was carried out by four different participants.

6.5.2 Participants

The participants consisted of 168 crowdsource workers from MTurk. At the end of the HIT, they were asked demographics questions whose answers are summarized in Table 6.3. These demographics are not fully consistent with some other previous studies on MTurk workers. For instance, the majority of our workers were males (59% in our experiment versus 25% [Paolacci et al., 2010] and 48% [Ross et al., 2010] in MTurk workers studies). Also, considering education and occupation in Table 6.3, our participants were considerably educated. This could be due to a self-selection bias (refer to the HIT title

and details in the next section). Five (3%) reported having colour blindness, but all of our visualizations used a colour-blind friendly palette from ColorBrewer [Harrower and Brewer, 2003] (see Section 6.4.2).

Table 6.3: *The demographics of the 168 participants.*

Gender	Female: 41%, Male: 59%
Age	Median: 29, Mean: 32, Range: [18,64]
Residence	USA: 47%, India: 40%, Other: 13%
Education	Bachelor's Degree: 45%, Some College, No Degree: 22%, Master's Degree: 15%, Other: 18%
Occupation	Professionals, Managers: 38%, Labourers, Service: 30%, Students: 18%, Unemployed, Retired, House-Makers: 15%
Colour-blind	None: 163, Red-green: 4, Other: 1

Percentages may not add up to 100 due to rounding

6.5.3 Procedure

We first conducted a pilot study with subjects in France and UK (N = 14; two participants per visualization type) by hosting the experiment form on a private web page. For the final experiment, HITs were uploaded on MTurk. Once all the required HITs were completed, the workers were granted a MTurk system qualification [Paolacci et al., 2010, p. 413] to carry out a follow-up questionnaire that was implemented in the form of another HIT.

Task

Participants had to fill out a form in their own Web browser. The form was split up into 10 pages and took around 25 minutes to complete. Participants could not review previous pages and could not proceed to the next page without completing all the questions. The first page instructed them to remain focused and not to stop unless all the pages were completed. The three problems were then presented on separate pages (one page per problem) either using text alone (V0) or text followed by a visualization (V1-V6). This is illustrated in Figure 6.6 for *Mam* and V4. Workers had to enter two values, v_1 out of v_2, and had to indicate their confidence on a 5-point Likert scale. The next page contained three catch questions in relation to the three previous problems, whose answer had to be selected from a drop-down list (Table 6.4). This was followed by four pages containing the objective numeracy test, the subjective numeracy test, and the paper folding test (tests discussed in Section 6.4.5). The final page was a brief questionnaire asking workers demographics-related information and the methods (e.g., mathematical theorems; diagrams drawn or used), material (e.g., pen and paper) or tools (e.g., calculator) they used to solve the problems. The time spent on each page was recorded. An example of an entire HIT (with all of its 10 pages) is available in Appendix B.2.1. In this example, the problems were ordered as *Mam*, *Cab*, *Eco* and each one was accompanied by visualization type V4.

Table 6.4: *The catch questions used in Experiment 1.*

Catch Question	Items in Drop-Down List (correct answer in red)
Which topic among the following was **not** part of the problems presented to you before?	Car accident, History course, Airplane crash, Cancer, Students, Red cab, Economics
The women were screened for skin cancer.	Yes, No
The witness identified the faulty cab as white.	True, False

Figure 6.6: *The mammography problem (Mam) presented using text and visualization type V4.* After solving the problem, the subjects had to indicate their confidence in their answer on the bottom Likert scale.

MTurk Design

Since standard MTurk markup does not support custom JavaScript, we used the external HIT hosting method. Each of the 42 unique combinations of problem orderings and visualization types (6 problem orderings × 7 visualization types) was a unique HIT on MTurk and four copies of each were uploaded. The title of the HIT was 'Scientific Study on Judgment and Visualization'. According to Horton et al.'s

[2011] reservation wage, workers are willing to carry out HITs for $1.38 per hour. So often, a better reward is granted (e.g., Paolacci et al.'s [2010] reward of $1.66 per hour). We opted for a reward of $1 for our 25-minute HITs (i.e., $2.40 per hour). A system qualification was also used to ensure that all the participating workers have a HIT approval rate (i.e., the number of approved HITs—HITs carried out successfully and paid for—as a percentage of the number of completed HITs) of at least 95%.

After completion, multiple HITs carried out by the same worker were rejected (i.e., not paid as stated on our instruction page) and discarded from analysis. HITs with a wrong answer to one or more of the catch questions were also rejected and discarded from analysis. HITs were reposted until 168 (i.e., 6 problem orderings × 7 visualization types × 4 assignments; thus, 24 different participants for each of the seven visualization types) valid HITs were obtained.

6.5.4 Hypotheses

Our hypotheses were:

- H1a. Vis types V1-V6 yield lower Error than type V0;
- H1b. Vis type V2 yields lower Error than type V1;
- H1c. The improvements observed for Vis types V3-V6 over type V0 will be the highest for subjects having low scores in the paper folding task (low Spat);
- H1d. The improvements observed for Vis type V2 and V4-V6 over type V0 will be the highest for subjects with high Spat.

6.5.5 Results

Biases in Answers

Bias captures the discrepancy between the subjects' answers to the problems and the exact answers. Figure 6.7 shows the distributions of Bias per Problem and Vis of the 24 subjects assigned to each Vis.
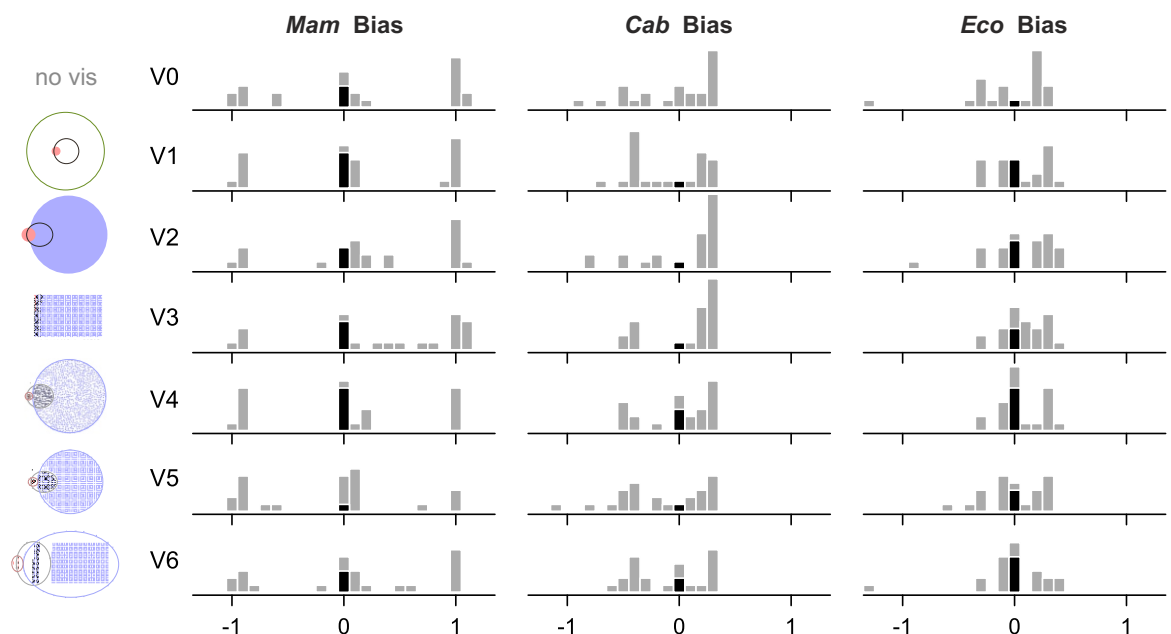
Figure 6.7: *Distributions of bias in answers per problem (column) and visualization type (row) with N = 24 each.* Black bars are exact answers. A bias of −1 is an answer 10× lower and bias 1 is 10× higher.

From Figure 6.7, it can be observed that: (i) consistent with previous findings [Gigerenzer and Hoffrage, 1995], answers are not normally distributed with certain wrong answers being much more common than others; (ii) the distributions differ across PROBLEMS but seem very similar across VIS types. In particular, there is no obvious sign of V1-V6 outperforming V0.

The distributions are also well-balanced around zero, suggesting no clear general tendency to underestimate or overestimate probabilities. Median biases per PROBLEM × VIS were overall close to zero (for all 21 values of median biases $M = 0.06$, $SD = 0.13$), as well as mean biases ($M = 0.003$, $SD = 0.14$). A grand mean of 0.003 is remarkably small (it would correspond to a probability estimate of, for instance, 0.1007 instead of 0.1) and suggests a 'wisdom of the crowd' effect [Surowiecki, 2005].

Exact Answers and Errors

Exact answers are those for which EX = 1 or equivalently BIAS = 0. They are shown in Figure 6.7 as black bars. There were 12% exact answers overall, with 15% exact answers for *Mam*, 5% for *Cab* and 15% for *Eco*. V0 yielded 6% exact answers, whereas V1-V6 yielded 14%, 11%, 11%, 21%, 7% and 14% exact answers respectively ($N = 24$ each). These percentages are much lower than in previous studies (e.g., overall exact answers: 12% in our study versus 40%-80% in previous studies as discussed in Section 2.5.5 and Section 3.9), suggesting that the problem does not lie in a systematic bias (e.g., no evidence for a base rate fallacy), but rather in poor individual accuracy.

A finer measure of accuracy is the ERROR metric, that indicates how far the subject's answer was from the exact answer. The overall median ERROR was 0.27 ($M = 0.38$, $SD = 0.39$). A mean of 0.38 is fairly large and corresponds to a probability estimate of 0.24 or 0.04 instead of 0.1.

Effect of Visualization on Error

In the rest of this section, we average the errors for *Mam*, *Cab* and *Eco* into a single ERROR measure. Box plots of ERROR per VIS are shown in Figure 6.8. It can be seen that effect sizes are quite small (differences in medians are much smaller than variances), and a Kruskal-Wallis rank sum test for non-normal distributions indeed reports no significant difference between VIS types ($H(6) = 4.1$, $p = 0.66$). Thus, our hypothesis H1b (i.e., V2 is more accurate than V1) is not confirmed.
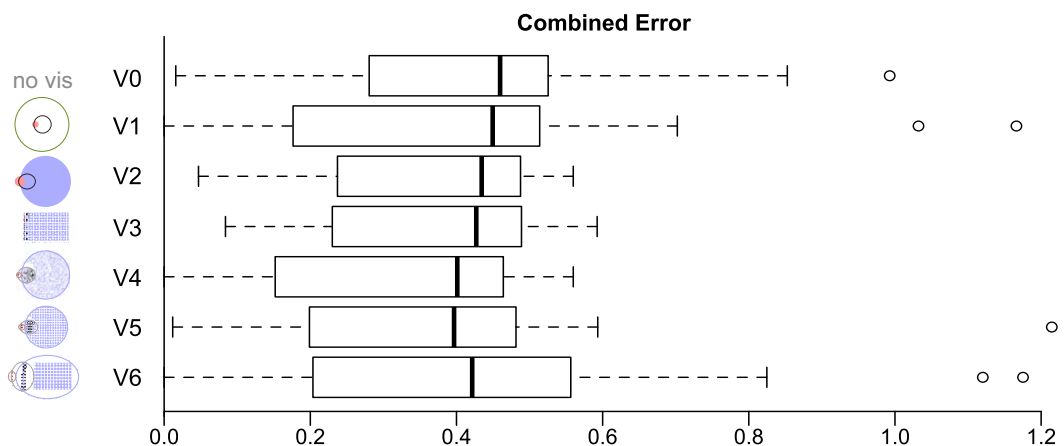
Figure 6.8: *Answer errors for the three Bayesian problems combined per visualization type (N = 24 each).*

Comparing text alone with V1-V6 aggregated ($N = 144$) does not yield any significant difference either (Kruskal-Wallis, $H(1) = 1.6$, $p = 0.21$). Thus, our hypothesis H1a (i.e., V1-V6 are more accurate than V0) cannot be confirmed either.

Subject Abilities

The median score for numeracy NUM was $24/30$ ($M = 23.5$, $SD = 4.0$), while that for spatial ability SPAT was $5/10$ ($M = 5.1$, $SD = 2.2$). More than half (55%) of the subjects got a moderate to high score for both tests, as defined by [Brown et al., 2011; Fagerlin et al., 2007b]. Correlations between ERROR and subject abilities were low ($r = -0.08$ for NUM and $r = -0.07$ for SPAT), suggesting little influence of abilities on accuracy overall.

We conducted further analysis by splitting subjects into two groups: SPAT ≤ 5 and SPAT > 5. In both groups, mean errors were very similar ($M \in [0.36, 0.38]$) for

- text alone with no visualization (V0),
- visualizations with glyphs (V3-V6) combined,
- visualizations with an Euler diagram (V2 and V4-V6) combined,

with no statistically significant difference. Thus, our hypotheses H1c (i.e., V3-V6 are more accurate than V0 for low Spat) and H1d (i.e., V2 and V4-V6 are more accurate than V0 for high Spat) on effects of individual abilities are not confirmed.

Confidence in Answers

The median score for CONF was 3 on a 5-point Likert scale ($M = 3.4$, $SD = 1.1$): subjects were typically 'reasonably confident' in their answers, with a trend towards high confidence. Correlation with ERROR was low ($r = -0.08$). Medians for CONF averaged across problems ranged from 3.0 to 3.7 depending on the VIS type, with no statistically significant difference (Kruskal-Wallis, $H(6) = 5.8$, $p = 0.45$).

Time Spent

The median completion time for reading and answering problems (TIME) was 113 seconds ($M = 154$ seconds, $SD = 130$): subjects typically spent about 2 minutes on each problem, which is about half the time spent by our initial pilot subjects (researchers and students from our laboratories). The correlation with ERROR was low ($r = 0.02$). The median TIME for the first problem presented was 129 seconds and went down to 106 seconds for the last one, suggesting a moderate learning effect.

An analysis of variance (ANOVA) with the model log(TIME) ~ VIS suggests that the total time spent on the three problems depends on the VIS type ($F(6, 161) = 2.5$, $p < .05$). Post-hoc pairwise comparisons with t-test and Bonferroni correction reveal a difference between V1 and V2 and between V2 and V4 ($p < .05$). Median times were 168 seconds for V1, 72 seconds for V2, and 149 seconds for V4. Thus, although subjects were not more accurate with V2 than V1 (our initial hypothesis H1b), they spent much less time (less than half) parsing V2.

6.5.6 Qualitative Feedback

When asked which tools or methods they used to get their answer, 8% of the participants specifically reported using Bayes' theorem (not mentioned in the question). Others used basic mathematics, carried out mental calculations, estimated answers, or guessed. About 75% reported using pen and paper or a calculator. Three subjects (2%) commented that they did not know they could use any of these tools.

Three weeks after the experiment, all participants were invited to complete a 2-minute follow-up questionnaire for $0.20 on MTurk (a preview of the HIT is available in Appendix B.2.2). Fifty-three participants out of the 168 (32%) responded. The seven visualization types for the *Cab* problem were shown and the participants had to either identify the type they had seen during the experiment or state that they do not remember at all. They then had to indicate on a 5-point Likert scale how much they looked at the provided visualizations, how much they used them to solve the corresponding problems, and explain why. This helped us understand whether the participants referred to the visualizations and whether they found them helpful.

A large majority (89%) of the 53 respondents reported using the visualizations to solve the problem (scales 3 'somehow used them' up to 5 'used them a lot' on a 5-point Likert scale). However, only 47% remembered the correct visualization type. This was likely due to the relatively long time period between the experiment and the questionnaire and due to the similarities among some visualization types. Among those who correctly remembered the visualization type, 92% reported using the visualization to solve the problem.

Most of the 53 respondents (79%) commented that the diagrams helped them visualize the problem (e.g., Table 6.5a), of which 6 (11%) specifically indicated that it helped them understand the relations between the sets and the given values (e.g., Table 6.5b). Four subjects (8%) who had text alone (V0) wished they had diagrams (e.g., Table 6.5c). Five (9%) reported comparing the colour and size of the sets and the overlapping regions, while 3 (6%) reported counting the glyphs. Other positive comments included: the diagrams helped to understand the statistics (e.g., Table 6.5d), clarified the problem (e.g., Table 6.5e), helped to identify the regions for the final answer (e.g., Table 6.5f).

Table 6.5: *A few of the participants' comments in the follow-up questionnaire.*

a	It can be explained only if there is a diagram relevant to the incident or question.
b	Diagrams show relationships between different sets. Easy to understand links.
c	I had trouble visualizing the data on my own. Diagrams would have been essential.
d	Diagrams gave me a visual clue to understand problem better and to find answer.
e	It would have been helpful to picture the scene.
f	Saw the diagram and figured out the regions for the answer.
g	I compared the size of the circles and the overlapping regions, but just looking at the diagram was not enough.
h	Used diagrams to get answer but did not fully understand them so it was difficult.
i	When thinking about populations, sample sizes and similar statistical problems, an intuitive grasp of problem comes quicker and easier through diagrams than mere words. I would not, however, have slavishly followed diagrams since I would still have to determine if the diagram was accurate and represented the problem as stated.
j	The text was enough, they didn't simplify anything, just complicated it more.
k	The text was more than sufficient.

However, 7 participants (13%) commented that the diagrams were not enough to answer the question (e.g., Table 6.5g). Three (6%) said that they did not fully understand the diagram (e.g., Table 6.5h), while one reported a lack of familiarity with the visualization. Six (11%) used the diagrams just to verify their answer, 4 (8%) stated that they did not trust the diagrams as in some experiments they are drawn incorrectly on purpose (e.g., Table 6.5i), while 3 (6%) claimed that having two representations was confusing and since the actual values were in the text, they ignored the visualization (e.g., Table 6.5j). Three (6%) stated that the diagrams were useless (e.g., Table 6.5k).

6.5.7 Discussion

Though we used the best known textual representation, that is, natural frequencies [Gigerenzer and Hoffrage, 1995], and the same text and text+visualization format as in previous studies, we failed to replicate previous findings as our subjects' accuracy was remarkably lower. For instance, considering V0 (text and no visualization), previous studies reported 26% [Garcia-Retamero and Hoffrage, 2013], 35% [Brase, 2009], 46% [Gigerenzer and Hoffrage, 1995], 51% [Sloman et al., 2003] and 72% [Cosmides and Tooby, 1996] exact answers with the natural frequency format, which are considerably higher than the 6% we obtained. Additionally, we could not confirm the effect of abilities and we could not determine which representation was most effective for different user types and level of abilities.

However, everything seems to suggest that our participants completed the tasks seriously. Only workers with a high HIT approval rate of at least 95% could participate, and since such a selection criteria is common practice, they have incentives to maintain a high rate. In addition, they were overall successful at the tests, the paper folding test being particularly attention-demanding. Finally, they often reported being confident in their answers and wrote a number of positive comments.

Of much concern is the fact that nearly all of our subjects were at least 'reasonably confident' in their answer. In a real life, fallacious reasoning combined with overconfidence can easily lead to wrong decisions, with potentially harmful consequences (e.g., undertaking chemotherapy after being diagnosed with cancer). Since we found that bias was close to zero overall, bad judgements could go either way.

We believe that far from invalidating the choice of the crowdsourcing method, these poor results motivate the need for more studies of this sort. It is important that techniques that aid probabilistic reasoning and decision-making can benefit untrained people from different backgrounds and age ranges and remain effective in situations where little time and attention are available.

In addition to a low accuracy overall, none of the visualizations we tested seemed to help. Although there is surely an effect [Cohen, 1994], Figure 6.8 suggests that effects on our metrics of interest (see Section 6.4.4) are small. This contrasts with conclusions drawn from studies on proportions of correct answers [Brase, 2009]. Girotto and Gonzalez [2001] go as far as arguing that with representations that clarify the subset relations of the problem and that use natural frequencies, anyone (including 'naive' subjects) could solve Bayesian problems, even if the subject is unfamiliar with Bayes' theorem or the representation. However, our results are also consistent with previous studies that reported no significant improvement with visualizations over text alone [Cosmides and Tooby, 1996].

Although Cosmides and Tooby suggest that passive visualizations are ignored when solving Bayesian problems [Cosmides and Tooby, 1996], 89% of our 53 participants who completed the follow-up questionnaire confirmed that they at least 'somehow' used the diagram. Most of them reported finding

the diagram very useful and commented that they generally like being given diagrams to solve problems. So it seems that subjects tend to overestimate the degree to which diagrams are helpful.

It is worth noting that despite positive comments overall, various participants commented that: they did not fully understand the diagram; having two representations was confusing; they ignored the diagram; the information provided in the text was sufficient. A few others doubted the credibility of the diagram and chose to trust the text. However, the answer to the Bayesian problem is in the diagram (as seen in Section 6.2) and so, subjects have to understand and trust the visualization for it to aid reasoning.

Several solutions have already been suggested. Teaching Bayesian reasoning is one of them. It can be extremely effective [Sedlmeier and Gigerenzer, 2001] but again, we chose to focus on techniques that can be used without prior training or background. Another well-known approach is the use of active rather than passive visualizations [Cosmides and Tooby, 1996] (but see also [Brase, 2009]). Having people draw their own visualization can be effective, but it is not practical in many situations, including in scientific press, information pamphlets or broadcasted visual adverts [Gigerenzer et al., 2007]. For similar reasons, static visualizations should probably be considered before interactive ones.

Macchi [2000] argued that Bayesian problems are difficult to solve because their presentation does not clarify the link between the numeric data and the sets and to which set overlap the value is referring to. So Kellen [2012] labelled all the set overlaps mentioned in the text and added the numeric value to the label (e.g., Figure 3.37). However, he reported better results when the problem was split up into two sub-problems (asking two questions instead of one and using two diagrams) and allowed subjects to move back and forth between the two (e.g., Figure 3.38; discussed in Section 3.9.1), which presentation requires more time, effort and possibly resources and is thus not practical in real-life.

Since the answer to the Bayesian problem is in the diagram itself, it should be possible to increase the chances that people find it, either by (i) helping them make the link between the text and the diagram (a few techniques in Section 2.6; further reading in Appendix C.3), (ii) encouraging them to search for the solution in the diagram, or (iii) forcing them to search for the solution in the diagram.

We propose two alternative presentation techniques that involve moderate modifications to the text:

- adding short instructions in the text that refer to the diagram—supporting (i) and possibly (ii);
- removing all numerical quantities from the text—supporting (iii).

The latter technique is based on the idea that a rough estimation can be better than a plain wrong calculation and should further eliminate any possible doubt that the visualization is not credible.

We conducted a second experiment in order to:

- confirm that simply adding a visualization to a Bayesian problem is of little help;
- testing whether an improvement can be obtained by the above two presentation techniques.

In order to test the former, we only included V0 and V4 and increased statistical power with larger sample sizes (from $N = 24$ to $N = 120$ each). V4 was the diagram with the lowest mean error and combined both Euler diagrams and glyphs. To further simplify the design, we only included the *Mam* problem. We chose it because it is a classic problem which has been used in various studies (e.g., [Garcia-Retamero and Hoffrage, 2013; Gigerenzer and Hoffrage, 1995; Hoffrage and Gigerenzer, 1998; Kellen, 2012; Mellers and McGraw, 1999; Sedlmeier and Gigerenzer, 2001]).

6.6 Experiment 2: Alternative Text Formats

In this second experiment, we aimed to (i) confirm with a larger sample that simply adding a visualization to a Bayesian problem is of little help, and (ii) investigate some possible solutions.

Based on the data from our first experiment, we hypothesized that simply appending a visualization to the textual information will yield at best weak improvements (our data only shows non-significant differences in mean Errros of about 0.1 points). We further hypothesized that this issue could be addressed by providing instructions in the text on how to parse the visualization. We also believed that the numerical values provided in the text could encourage wrong calculations and discourage parsing the visualization, so we hypothesized that not providing these values would also help.

6.6.1 Design

Our independent variable was presentation type Pres, with the following four conditions:

- V0: Classic text with no visualization, as condition {Problem=*Mam*, Vis=V0} in Experiment 1;
- V4: Classic text with visualization, as condition {Problem = *Mam*, Vis = V4} in Experiment 1;
- V4a: LikeV4, but with references to the visualization in the text (Table 6.6);
- V4b: LikeV4, but with numbers removed from the text (Table 6.6).

We used a between-subjects design and our dependent variables were Bias, Error, Ex, Time and Conf as in Experiment 1.

Table 6.6: *Novel text format for the Mam problem in Experiment 2.*

V4a	10 out of every 1,000 women at age forty who participate in routine screening have breast cancer (*compare the red dots in the diagram below with the total number of dots*). 8 of every 10 women with breast cancer will get a positive mammography (*compare the red dots that have a black border with the total number of red dots*). 95 out of every 990 women without breast cancer will also get a positive mammography (*compare the blue dots that have a black border with the total number of blue dots*).
V4b	A small minority of women at age forty who participate in routine screening have breast cancer. A large proportion of women with breast cancer will get a positive mammography. A small proportion of women without breast cancer will also get a positive mammography.

6.6.2 Participants

Our participants were 480 MTurk workers who never completed any of our previous HITs. We thought the demographics should be similar to Experiment 1 (details in Section 6.5.2), so we did not include any demographics questions in our HITs.

6.6.3 Procedure

We conducted the experiment on MTurk, as Experiment 1. The HITs were shorter, including only the *Mam* problem and no abilities tests.

Task

The form was made up of four pages and took around five minutes to complete. After the instruction page, the *Mam* problem was presented using either V0, V4, V4a or V4b. The questions and the layout of the pages were the same as before (see Figure 6.6), except participants were told they could optionally use any tool or method. The next page had a single catch question (similar to the first catch question in Experiment 1; see Table 6.4). On the final page, participants were asked whether they tried to compute an exact answer and had to indicate on two independent 5-point Likert scales how much they used the information in the text and the diagram to solve the problem.

An example of an entire HIT (with all of its four pages) is available in Appendix B.2.3. In this example, the *Mam* problem is presented using presentation type V4b (i.e., text with no numbers).

MTurk Design

We created four unique HITs, one for each presentation type. For each unique HIT, 120 assignments were uploaded on MTurk with a $0.40 reward. The same system qualification as in Experiment 1 was used. Participants to Experiment 1 were blocked and duplicate HITs were rejected as before. HITs of workers who selected the wrong answer to the catch question were also rejected. In contrast to the previous experiment, workers who submitted the HIT only once but had previously seen the problem (either because they gave up before submitting or experienced technical issues), were identified through a question on the last page asking the workers whether they already attempted to load the HIT and failed to submit it. We clearly stated that this was not going to affect our decision to accept or reject the HIT. In fact, HITs from workers who replied "yes" or "unsure" (about 11%, not included in $N = 480$) were accepted (i.e., paid) and the data was discarded from analysis.

6.6.4 Hypotheses

Our hypotheses were:

- H2a. Condition V4 does not lower the mean ERROR by more than 0.1 points compared to V0;
- H2b. Condition V4a yields lower ERRORS than V4;
- H2c. Condition V4b yields lower ERRORS than V4.

6.6.5 Results

Bias

Figure 6.9 shows the distributions of BIAS for each PRES. These are similar to those of the previous experiment, except for V4b which is closer to normal. This time, as shown in Figure 6.10, not all the distributions were well-balanced around zero. The median biases were 0.015, -0.59, -0.49 and 0.013 for V0, V4, V4a and V4b. The differences are statistically significant (Kruskal-Wallis, $H(3) = 22$, $p < .001$).
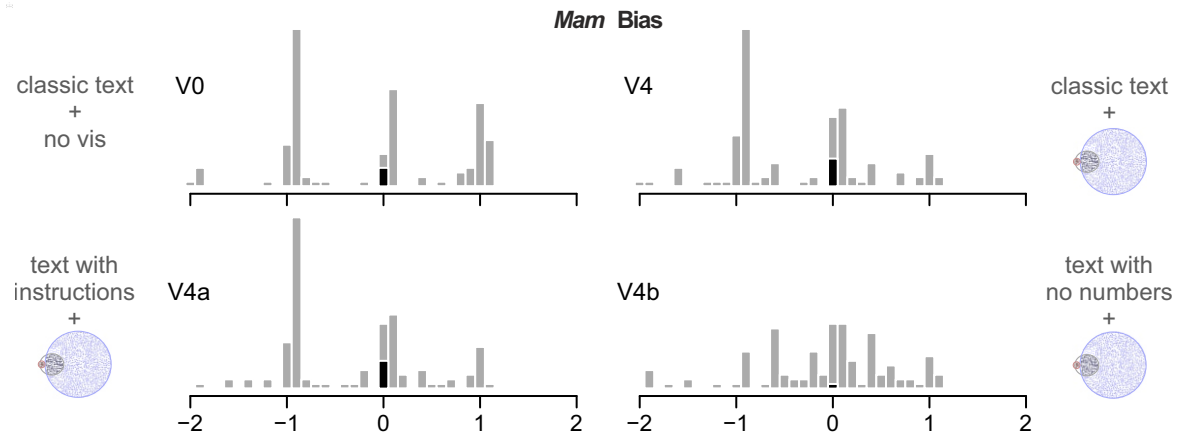
Figure 6.9: *Distributions of biases in answers to the Mam problem (N = 120 each) per presentation type.* Black bars are exact answers.
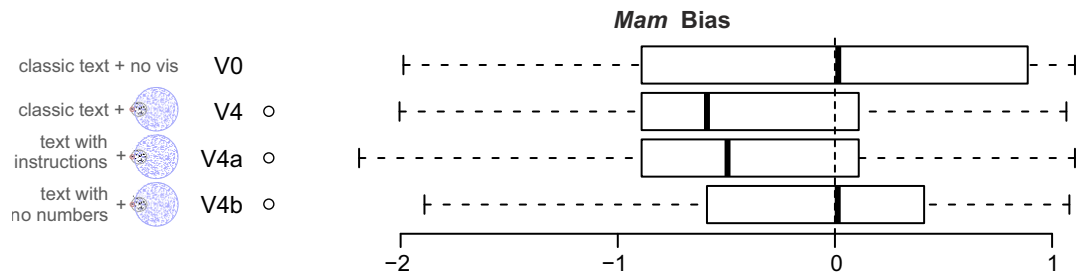
Figure 6.10: *Biases in answers to the Mam problem per presentation type (N = 120 each).*

Accuracy

Exact answers for V0 and V4 were respectively 3.3% and 5.0% (see black bars in Figure 6.9) and lower than in Experiment 1 (15% for *Mam*). Median errors for V0 and V4 were both 0.8903 and equal to those in Experiment 1. Mean errors were however larger (0.76 and 0.68 compared with 0.64 and 0.54 in Experiment 1). Thus, it seems that subjects were overall less successful in this experiment.

That V4 yielded a lower mean error than V0 (0.68 versus 0.76, with same medians) suggests that adding a diagram might have helped, but the difference in means is consistent with H2a. Inconsistent with H2b however, V4a yielded the same ratio of exact answers, the same median error and a similar mean error as V4, suggesting that referring to the diagram in the text did not help. In contrast, although V4b rarely yielded an exact answer (0.83%—a single subject out of 120 gave an exact answer; later, in the questionnaire, this subject commented, "I kept losing count, so I hit PrintScreen, and pasted it into paint, and marked the ones I had counted."), it yielded a much lower median error of 0.41 and a lower mean error of 0.53, consistent with H2c.

The differences between PRES types, shown in Figure 6.11, are statistically significant (Kruskal-Wallis, $H(3) = 21$, $p < .001$). A multiple comparison with V0 using Siegel and Castellan's procedure revealed that all visual presentation types are significantly better than text alone, confirming our previous hypothesis H1a (R package kruskalmc, one-tailed, $p < .05$). Using the same test with V4 as the control, the only significant difference is with V4b, confirming our hypothesis H2c (removing numbers helps) but not H2b (referring to the figure helps). Finally, a Wilcoxon test of equivalence for non-normal data confirms our hypothesis H2a (R package etc.diff, margins = [–0.1, 0.1], $p < .05$).
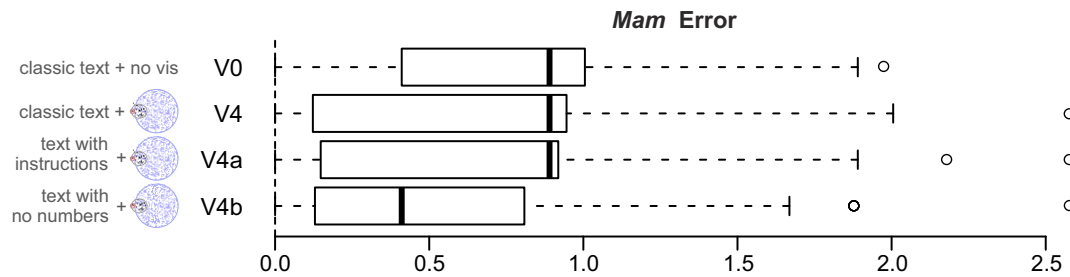
Figure 6.11: *Answer errors to the Mam problem per presentation type (N = 120 each).*

Hence, all visual presentations were better than text alone but improvements were small, except for V4b where improvements were larger. The task involved in V4b was still error-prone, with probability estimations typically about 3 times lower or higher, but it clearly improved over V0 for which typical estimates were 6 or 8 times lower or higher (depending on whether we consider the average or the median error). In addition, V4b yielded no bias, whereas with other visual presentations, subjects tended to underestimate probabilities.

Confidence

Confidence scores were similar to Experiment 1 and very similar across PRES, with a median of 3 ('reasonably confident') and means between 3.28 and 3.36. Differences were not statistically significant (Kruskal- Wallis, $H(3) = 0.6$, $p = 0.89$). As before, correlation with error was low ($r = -0.06$).

Time

Completion times were similar to those of Experiment 1 and similar across all PRES conditions (ANOVA, $F(3, 476) = 1.26$, $p = 0.29$). Medians ranged from 112 seconds for V4b ($M = 145$, $SD = 115$) to 132 seconds for V4a ($M = 163$, $SD = 108$).

Strategies

Discarding 'unsure' responses (9% overall), subjects who reported attempting to get the exact answer were 68%, 72%, 63% and 50% for V0, V4, V4a and V4b respectively. This suggests that the diagrams did not dissuade subjects from trying to find the exact answer.

As for the degree to which subjects reported relying on the diagram, the median answer to the 5-point Likert scale was 3 for V4 ($M = 3.3$, $SD = 1.5$). Subjects who were assigned V0 and later asked whether they would have used the diagram gave a median answer of 4 ($M = 3.6$, $SD = 1.$), suggesting that subjects tend to use diagrams less than they would have predicted.

V4a was similar to V4, with a median answer of 3.5 ($M = 3.4$, $SD = 1.3$). In contrast, for V4b, the median answer was 5 ($M = 4.4$, $SD = 0.86$). This indicates that, rather unsurprisingly, subjects relied on the diagram much more when numbers were not provided.

6.6.6 Discussion

Our first experiment revealed that crowdsource workers were quite unsuccessful at solving Bayesian problems. Our second experiment had two objectives:

1. confirm that simply adding a visualization (consisting of an Euler diagram or glyphs or both) to a Bayesian problem is not a viable solution (we increased sample size from $N = 24$ to $N = 120$ to get more statistical power);

2. start exploring possible solutions to this problem.

We met our first objective by measuring a statistically significant difference between text alone and text+diagram, but showing that the practical difference was small (no more than 0.1 points of mean error as confirmed by an equivalence test).

We incidentally found that subjects were even less accurate than in Experiment 1, although the reasons for this are unclear. The design differed in two respects: (i) subjects were initially instructed they could use any tool but they did not have to; (ii) those who reported having previously seen the HIT without submitting were discarded from the analysis (11%). So subjects might have been primed in trying to get exact answers and failed in their calculations, or alternatively, our first experiment could have overestimated subjects' accuracy by including workers who possibly acquainted themselves with the problem way before carrying out the task.

With respect to our second objective, we tried two alternative presentations and found that referring to the diagram within the text (V4a) did not help, whereas removing the actual numbers from the text (V4b) yielded clear improvements. This is an important finding, since it is a simple and effective technique that can be easily applied to many real-life situations. Previously proposed techniques are more difficult to apply since they either involve prior training [Sedlmeier and Gigerenzer, 2001] or, like active constructions, require a pen and paper, time and possibly assistance [Cosmides and Tooby, 1996] (also [Cox, 1999]). Though Kellen [2012] removed the numeric data from the text, he still included the values as labels in the diagram (as shown in Figure 3.37; also, Kellen's work was published after ours [Micallef et al., 2012]). Thus, we do not know of any previous study on Bayesian reasoning which proposed a technique similar to V4b (a presentation with no numbers, be it in the text or the labels in the diagram), possibly due to the fact that most of them focused on increasing the number of exact answers rather than reducing estimation errors.

Our particular study focused on *how* to reduce inaccuracy, excluding investigations on *why* exactly subjects make mistakes and what these mistakes are. Some previous studies have observed different types of miscalculations in Bayesian reasoning tasks [Gigerenzer and Hoffrage, 1995] and such investigations are likely crucial for designing effective visualizations and text formats. The distributions in Figure 6.9 do suggest that typical miscalculations or reasoning errors have occurred in our experiments. Some of them (close to bias 1) seem less common when a visualization is provided, possibly resulting in more miscalculations of a different type (close to bias -1) and a lower bias overall (Figure 6.10). Trying to address specific mistakes might help to further increase accuracy, including when no numbers are provided, since diagrams can be misinterpreted. However, recurrent mistakes seem to vary across problems (Figure 6.7) and finding general solutions might be challenging.

Our findings on visualization-friendly textual formats are only a step towards helping people being more accurate in Bayesian reasoning: even when shown a diagram without numbers, workers were still inaccurate at estimating probabilities (with a typical estimation about 3 times lower or higher). There

are yet further possible improvements to our presentation techniques. Our working memory capacity [Cowan et al., 2005] and our capabilities to process relations [Kroger et al., 2002] should also be taken into account, as a text+visualization presentation that seemly links the various elements in the problem should reduce the required cognitive effort to understand and solve a problem [Cook, 2006].

We will now discuss possible improvements to our presentation techniques and other future work.

6.7 Future Work

Our results indicate that only adding a visualization to the classic text is not enough. So various techniques to effectively combine visualization and text for Bayesian reasoning should be evaluated. A few classic techniques have been discussed in Section 2.6 and others are mentioned below (in the next two paragraphs), but further reading is provided in Appendix C.3.

A possible improvement to our presentation techniques is to show the numbers on the visualization (similar to that proposed by Kellen [2012]—Figure 3.37). Tufte claims that, "Words on and around graphics are highly effective" [Tufte, 1983, p. 182] and advises to "Write out explanations of the data on the graphic itself" [Tufte, 1983, p. 56] (Tufte's second graphical integrity principle). So, the text (story) could be overlaid on the visualization. Alternatively, miniature visualizations like Tufte's [2006] sparklines (Appendix C.3.3) or as in Figure 2.11C in Section 2.6 could be embedded in the text.

Storytelling techniques (e.g., [Kosara and Mackinlay, 2013; Segel and Heer, 2010]; Appendix C.3.6) could also be adopted. Tufte demonstrated the benefits of small multiples (i.e., same structured, small visual displays) to illustrate how a concept changes as a variable changes [Tufte, 1990, pp. 67-79] (Appendix C.3.1). Ware argued that it is easier to convey a story through strip cartoons and diagrammatic instructions (e.g., [Heiser et al., 2004]) than using just pictures [Ware, 2008, p. 145] (Appendix C.3.4). Studies in journalism and mass communication suggest that the most effective way to present information related to a story is to use text and graphics (e.g., [Huh, 1993; Pasternack and Utt, 1990]) and to provide the information of the graphic also in the text (e.g., [Griffin and Stevenson, 1992, 1994b, b, 1996]) (Appendix C.3.5). Thus, it could be helpful to present the Bayesian problem as small multiples such that small visual displays, each with some text and a relevant diagram, are provided in sequence as in strip cartoons or diagrammatic instructions to outline the story of the Bayesian problem. This could be similar to our visual explanation of the classic mammography problem in Section 6.2.

A story provides a context to the visualization, which will help the reader understand the problem, the provided information and the data that is required to solve the problem [Shapiro, 2004]. There must be a sense of trust between the storyteller and the reader [Glassner, 2004, p. 93] and so, the problem should be realistic (e.g., avoid using problems about 'a Disease X'), the text should be engaging and the story should be one that the reader can relate to and imagine. Thus, considering for instance the classic mammography problem, it would be interesting to compare the accuracy of women versus men and that of women who once had cancer or experienced cancer through a relative or friend versus others who never experienced cancer. It would also be interesting to evaluate a personalized presentation that uses the pronoun 'you' rather than a noun, for instance, 'a woman'. These findings could then be compared to those of another presentation where the subjects have to imagine they are advising or choosing for someone else (e.g., they are a doctor or a parent) rather than for themselves (e.g., they are a patient). The latter presentation is interesting as a study indicates that roles affect the decision-making process and the preferred treatment in a medical situation (e.g., subjects assigned the role of a doctor had a

stronger preference than those assigned the role of a patient for a preventative treatment to reduce the chances of dying from a disease rather than no preventative treatment and an increased risk of dying from the disease) [Zikmund-Fisher et al., 2006].

Changing the order of the different pieces of information presented in the classic text could reduce the likelihood that the base rate is ignored or underestimated, due to the recency effect as subjects are more likely to rely on details presented at the end of the problem [Krosnick et al., 1990] (also [Obrecht and Chesney, 2013]). Thus, the ordering of information in the text could be worth further investigation. In fact, more novel textual representations with no visualizations should be proposed and evaluated, as visualizations are not helpful for all, particularly for those individuals who do not have the right levels of ability to handle graphical information [Gaissmaier et al., 2012].

Although we focused on static representations due to their wider applicability, interactive techniques can not only help users understand the relationship between the text and the diagram (e.g., [Tsai, 2012] and [Spiegelhalter et al., 2011]—movie S3) and experience the inferencing process [Sedlmeier, 1999], but also let them explore variants and generalize the problems [Victor, 2011].

Other future work include extending our design space to evaluate other visualization types. Although our initial comparison was inconclusive, we anticipate that using problem formats that are more adapted to visualizations will encourage subjects to actually use them and ultimately make these visualization designs easier to compare in terms of both speed and estimation accuracy.

Other visualization designs that we could evaluate for Bayesian reasoning, include: Euler diagrams with equally-sized glyphs that are laid out as one block per region (Figure 6.12A) or multiple blocks per region with a size that is within a subitizing range (i.e., the number of objects that humans can effortlessly and correctly judge their number without counting) so that the number of glyphs is easily and accurately estimated without the need of focused attention [Gobet and Clarkson, 2004] (Figure 6.12B); designs that adopt alternative ways of representing quantitative data with glyphs in Euler diagrams (a few discussed in Section 3.8); an Euler diagram (e.g., V2) and a frequency grid (e.g., V3) as two adjacent diagrams in one display for one problem, so while the Euler diagram depicts only the set relations, the frequency grid depicts the quantitative data; a frequency grid with a tree diagram, as proposed but not evaluated by Spiegelhalter et al. [2011] (Appendix C.2.5) and other combinations of visualization types. Individual differences (e.g., background, age, abilities) among subjects should also be taken into account as they can affect diagram understanding [Ziemkiewicz and Kosara, 2009] and accuracy in solving the problem [Garcia-Retamero and Hoffrage, 2013] (more in Appendix C.2.9).
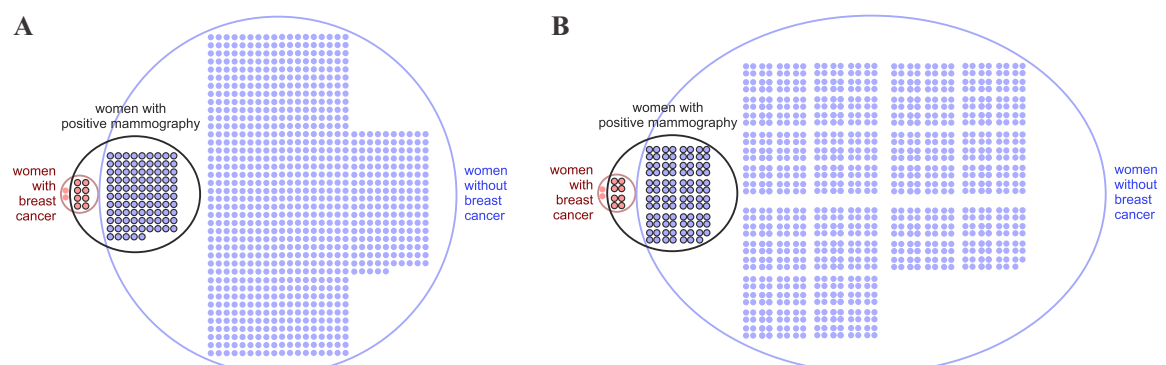
Figure 6.12: *Two visualization designs that we could evaluate in the future.*

In future studies, we could prohibit the use of pen and paper and any other external tools, except for a calculator (as done in Kellen's [Kellen, 2012] study). Doing so, could ensure that any positive effect is strictly accountable to the evaluated presentation type, without any addition aid from external tools that the subject could have used. Also, when using pen and paper as external tools, cognitive load increases as the subject has to keep track and link information in multiple locations [Pirolli and Card, 2005; Sweller, 1994], that is, the computer display and paper. We could also design a sophisticated incentive scheme [Shaw et al., 2011] to motivate the majority of the MTurk workers to carry out the HITs rapidly, yet accurately, as would be required in a real-life situation. Special attention should also be given to workers in India and any others who have difficulties with the language of the HIT, as they might not fully understand the instructions and the problem and thus, perform poorly [Shaw et al., 2011]. In such cases, it is likely that the problem is the language rather than the presentation type. To identify whether the language is an issue, 'instructional manipulation checks' [Oppenheimer et al., 2009] could be used.

6.8 Summary

We used crowdsourcing to assess the effect of six visualizations (based on Euler diagrams, glyphs and combinations of both) and text alone on probabilistic reasoning using three classic Bayesian problems in psychology. Our findings were inconsistent with previous studies in that subjects' accuracy was remarkably low and did not significantly improve when a visualization was provided with the text. A follow-up experiment confirmed that simply adding a visualization to a textual Bayesian problem is of little help for crowdsource workers. It however revealed that communicating statistical information with a diagram, giving no numbers and using text to just set the scene significantly reduces probability estimation errors. Thus, novel representations that holistically combine text and visualizations and that promote the use of estimation rather than calculation need to be investigated.

We also argue for the need to carry out more studies in settings that better capture real-life rapid decision-making than laboratories. We propose the use of crowdsourcing to partly address this concern, as crowdsourcing captures a more diverse and less intensely focused population than university students. Doing so, we hope that appropriate representations that facilitate reasoning for both laymen and professionals, independent of their background, knowledge, abilities and age will be identified. By effectively communicating statistical and probabilistic information, physicians will interpret diagnostic results more adequately, patients will take more informed decisions when choosing medical treatments, juries will convict criminals and acquit innocent defendants more reliably, intrusion alarms will be regarded and criminal attacks will be predicted and restrained.

Following our findings of no statistically significant differences between visualization types, even for Bayesian reasoning, it is still unclear whether Euler diagrams with glyphs are more effective than area-proportional Euler diagrams in representing the cardinality of the sets and their overlaps. Thus, further studies should be conducted to investigate whether there exists a visualization that avoids area judgement and that is more effective than area-proportional Euler diagrams in representing data set relations and relevant quantitative data.

We will now proceed to the concluding chapter.

Chapter 7

Conclusions

This chapter summarizes the objectives, contributions and findings of this dissertation and highlights important future work.

7.1 Objectives, Contributions and Findings

Through the exploration of different drawing techniques, curve representations and graphical encodings of quantitative data, we have developed novel drawing algorithms for Euler diagrams and we have conducted a user study to assess the effectiveness of Euler diagrams for probabilistic judgement.

The main drawing methods are eulerForce (Chapter 4), which uses a force-directed approach to lay out Euler diagrams, and eulerAPE (Chapter 5), which uses ellipses to draw accurate area-proportional Venn diagrams with three smooth curves. For the user study (Chapter 6), the effectiveness of area-proportional Euler diagrams, glyph representations, Euler diagrams with glyphs and text+visualization formats have been evaluated for Bayesian reasoning, and a method eulerGlyphs has been devised to automatically and accurately draw the assessed visualizations for any Bayesian problem. Figure 7.1 illustrates examples of diagrams drawn by our methods. This work has been published (Section 1.3). Software implementations and experimental stimuli and data are available online (Section 1.4).

We will now discuss details of our contributions and findings with respect to our research objectives.
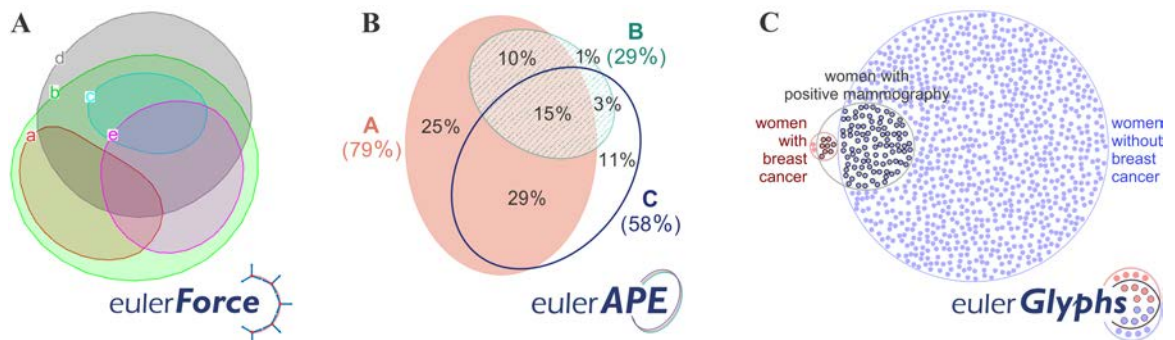
Figure 7.1: *Examples of diagrams generated by our methods: (A) eulerForce, (B) eulerAPE, (C) eulerGlyphs.*

Aesthetics and Perceptual Factors

A distinctive feature of Euler diagrams is the use of closed curves to depict the data sets. With such curves, elements within an enclosed region tend to be grouped together due to the perceptual grouping principles of common regions and closure (Section 3.1). Data set relations are easily noticeable when the curves (Section 3.5): have good continuity, so they pop out; are convex, so complete objects are perceived even when partially occluded; are regular, so they are easily identified. Wellformedness properties (Section 3.4.4) are also important for diagram comprehension (Section 3.5.1, Section 3.7.3). So our drawing methods generate diagrams with curves that have good continuity and if possible, with curves that are convex and circle-like and with all wellformedness properties satisfied, as in Figure 7.1.

There is a lack of empirical evidence to indicate what aesthetic features aid diagram comprehension and how the diagrams and their areas are perceived (Section 3.5, Section 3.7.3). Aesthetic criteria and measures need to be formalized. However, as in graph drawing (Appendix C.1.1), theories of perception and cognition have to be applied and user studies have to be conducted for such metrics to be fully formalized and evaluated. In this dissertation, we started work in this area by applying principles and theories of human perception and cognition, graphical excellence and integrity, graphical encoding, graphical perception, to Venn and Euler diagrams (Section 3.5, Section 3.7.3). We also prioritized some features related to curve aesthetics (e.g., the importance of closure and continuity followed by convexity and shape) (Section 3.5.2) and we used these as guidelines when devising our drawing methods.

Based on such theories of human perception, cognition and attention, designs that effectively use colours, textures and outlines, as in Figure 7.1B, were also investigated and adopted (Section 5.8). With the design of Figure 7.1B, the curves in which any point in the diagram is located are easily identified. Also, by attending to the feature type of a particular curve, a viewer can easily focus on that curve, as curves with other feature types recede. This can be particularly helpful for area-proportional diagrams where the curves depict set relations as well as relevant quantitative data, thus its use in eulerAPE.

Laying Out Euler Diagrams Using a Force-Directed Approach

We devised eulerForce to improve the layout of Euler diagrams generated by other methods (Chapter 4).

Our method eulerForce is novel as it uses a force-directed approach to lay out Euler diagrams. Being the first of this kind, we devised a simple yet an appropriate force model to improve the layout of incomprehensible diagrams generated by current drawing methods. This force model ensures that the diagram only depicts the required set relations as in the original diagram and strives to obtain curves that are regular, smooth, convex, similarly shaped and not too close to one another, zones that are connected and adequately sized, and contained curves that are centred in their containing curve or zone. Thus, it aims at improving important aesthetic features that could otherwise impede understanding (Section 3.5) and yet, only depict the required set relations as in the original, incomprehensible diagram.

Our evaluation of eulerForce indicates a great potential for such force-directed techniques. Accurate layouts with smooth curves were generated in relatively fast time for a large majority of the Euler diagrams with three to five curves that were automatically generated by Rodgers et al.'s [2008b] method (Section 4.4.1). Comparing eulerForce's layouts with those of the previous multi-criteria optimization method [Flower et al., 2003b], we noted that eulerForce's layouts were more compact, symmetric, easier to understand and their curves were more regular and similar in shape (Section 4.4.2), as in Figure 7.1A.

Drawing Area-Proportional Venn Diagrams Using Ellipses

We developed euler*APE* to automatically draw accurate area-proportional Venn diagrams with smooth curves for most data with three sets (Chapter 5).

Our method euler*APE* is the first to use ellipses and thus, the first to generate diagrams that are: easy to comprehend like those with circles but not polygons; accurate for most 3-set data like those with polygons but not circles. It is a novel, deterministic drawing method, which is neither an extension nor an adaptation of previous techniques. All the generated diagrams are also wellformed, as in Figure 7.1B.

We focused on Venn diagrams with three curves for mainly the following: 3-Venn diagrams are widely used; current methods having difficulties drawing such diagrams accurately with smooth curves; this work is the first to use ellipses. We extensively evaluated the drawing method and the effectiveness of ellipses to facilitate future extensions of this work. Some Euler diagrams can be drawn with euler*APE*, but further investigations are required to determine when accurate diagrams can be drawn with ellipses.

Our drawing method euler*APE* rarely encounters a local minimum and when it does, it effectively handles such cases to obtain a good solution whenever an accurate area-proportional 3-Venn diagram drawn with ellipses is known to exist for the given area specification (Section 5.7.1). A number of techniques have been devised to avoid and handle local minima (Section 5.6.7): (i) a novel cost function (Section 5.6.4); (ii) a novel rational starting diagram generator (Section 5.6.5); (iii) a mechanism to adjust the ellipses' properties during the optimization process (Section 5.6.6) using (iii-a) a cooling schedule to explore major changes in the diagram at the initial stages and minor changes in later stages of refinement, (iii-b) a rerun option which when enabled, reruns the optimization using larger starting values for its parameters that determine changes to the diagram. The various design options for each of these components were thoroughly evaluated before the most effective were employed by euler*APE*. The rationale for our choices was clearly explained and justified by experimental and pragmatic evidence.

Our evaluation (Section 5.7.1) suggested that a large majority of the diagrams (97.7%, N=10000) are drawn in a time (1 second) that ensures users' train of thought is maintained, and nearly all the diagrams (99.7%, N=10000) are drawn in a time (10 seconds) that ensures the users' attention is maintained.

Since euler*APE* generates an accurate diagram with respect to a given area specification whenever one is known to exist for the area specification (Section 5.7.1), we used euler*APE* to evaluate the effectiveness of ellipses in drawing wellformed accurate area-proportional 3-Venn diagrams for any 3-set data (Section 5.7.2). Our evaluation revealed that accurate diagrams can be generated for a large majority of random area specifications (86.1%, $N = 10000$) using ellipses, but for none using circles. There are area specifications for which an area-proportional wellformed 3-Venn diagram cannot be drawn accurately using convex curves (Section 3.7.5) and thus, such a result indicates great potential for using curves that are smooth like circles, but more general and flexible like ellipses. We also identified characteristics of area specifications for which (a) an accurate diagram was drawn with ellipses (Section 5.7.2) and (b) a close to accurate diagram with circles exists (Section 5.7.3).

We compared the effectiveness of euler*APE* with that of `venneuler` [Wilkinson, 2012]—currently considered the most effective method in drawing accurate diagrams with circles (Section 5.7.3). None of `venneuler`'s diagrams for the random 10,000 area specifications were accurate and only 64.5% were 3-Venn diagrams (all area specifications corresponded to the abstract description of a 3-Venn diagram). This indicates the limitations of circles in drawing accurate diagrams for most 3-set data.

We also compared diagrams generated by eight circle-based drawing methods, six polygon-based methods and eulerAPE for real medical data (Section 5.7.4). We noted that the diagrams with circles were easier to comprehend than those with polygons, but all of them were misleading and inaccurate. Half of the diagrams with polygons were accurate and most of the inaccurate diagrams were not as misleading as those with circles. In contrast, eulerAPE's diagram with ellipses (Figure 7.1B) was accurate and the curves were easily distinguishable. This indicates that polygons should be avoided when possible and instead, curves that have more degrees of freedom than circles but are similarly smooth should be used.

Assessing the Effect of Euler Diagrams with Glyphs on Bayesian Reasoning

We conducted a user study to assess the effect of different area-proportional Euler diagrams, glyph representations, Euler diagrams with glyphs and also text+visualization formats on Bayesian reasoning, (Chapter 6). We also devised the drawing method eulerGlyphs to automatically and accurately draw the assessed visualizations for any Bayesian problem.

Our user study is the first to assess the effect of six different visualizations over three Bayesian problems and on a large, diverse population through crowdsourcing. All previous studies were carried out in a laboratory. The visualizations consisted of two different area-proportional Euler diagrams, a glyph representation, and three different Euler diagrams with glyphs (Figure 7.1C represents one type) that have not been evaluated before our study. In contrast to previous studies, the diagram design rationale was clearly explained (Section 6.4.2) and the diagrams were automatically generated by our method eulerGlyphs (Section 6.3) to ensure an accurate depiction of the quantitative data and a consistent design (e.g., labels, colours, glyph size) between visualization types. Though eulerGlyphs was specifically devised to generate the diagrams for this study, it can be used to generate such diagrams for any Bayesian problem. It is also the first of its kind. In our study, we also proposed and evaluated two novel textual problem formats, designed to be used with visualizations to holistically combine the two. Our study is thus also the first to the evaluate different visual and textual representations using the "same experimental methodology" [Brase et al., 2006, p. 974]. To our knowledge, this is the only computer science study on Bayesian reasoning since Cole's work [Cole and Davidson, 1989; Cole, 1989].

Our study failed to replicate previous findings in that subjects' accuracy was remarkably lower and visualizations exhibited no measurable benefit, as their accuracy was not improved significantly when the visualization was provided with the text. A second experiment confirmed that simply adding a diagram to the textual problem is of little help, even when the text refers to the diagram. However, it suggested that visualizations are more effective when the text is given without numerical values and this information is conveyed through phrases like "large proportion", "small minority" that just set the scene. With such a text+visualization presentation, probability estimation errors in our experiments diminished significantly, indicating the need for numerical data to be presented in a way that estimation rather than calculation is encouraged and that the text and visualization are perceived holistically, as one.

Inconsistencies between our findings and those of previous studies could be due to differences in the population types and the experiment setting. We argue for the need of more studies to be carried out on heterogeneous populations of non-experts and in settings that better capture real-life rapid decision-making than laboratories. We recommend the use of crowdsourcing due to the diversity of the workers and the setting where the workers are less intensely focused than university students. Such studies will aid in identifying representations that are effective for both laymen and professionals.

Since the differences between visualization types with respect to the effect of visualization on error were statistically insignificant, it is still unclear whether Euler diagrams with glyphs are more effective than area-proportional Euler diagrams and thus, whether depicting quantitative data using a number of equally-sized discrete objects is more helpful that using area.

7.2 Future Work

Future work in relation to the aesthetics of Venn and Euler diagrams (Section 3.5, Section 3.7.3), our drawing algorithms, *eulerForce* (Chapter 4) and *eulerAPE* (Chapter 5), and our user study on Bayesian reasoning and the related drawing method *eulerGlyphs* (Chapter 6), have been discussed in detail in each of the respective chapters, namely in Section 3.5.4 and Section 3.7.3, Section 4.5, Section 5.9 and Section 6.7. In this section, we highlight important future work for each one.

In Section 3.5 and Section 3.7.3, we have applied principles of human perception and cognition to the aesthetics of Euler diagrams to explore design features that could easily convey set relations and their cardinalities. However, various user studies have to be conducted for aesthetic criteria and metrics to be formalized and evaluated in various contexts and application areas. At present, there is a lack of empirical evidence to indicate, for instance: the effect of various aesthetic features on diagram comprehension; how the diagrams are processed perceptually and cognitively; how zone areas are perceived and judged; the features that could affect area judgement; discrepancies in areas that are not noticeable. With such criteria, metrics and empirical findings, the effectiveness of layouts could be evaluated, forces would be added to *eulerForce*'s force model to handle each criteria and generate improved layouts, and a variant of *eulerAPE* would be devised to generate diagrams with important aesthetic features and with zone area inaccuracies that are not noticeable to human users. Diagram design features (e.g., colour, texture, outline) should be evaluated empirically for curves and zones to be easily identified and to allow users to attend to specific aspects of the diagram without much effort. Details about future work in this area are provided in Section 3.5.4 (also Section 3.5, Section 3.7.3).

In Chapter 4, we have demonstrated the effectiveness of a force-directed approach in laying out accurate diagrams with smooth curves in relatively fast time. To facilitate the development of the first force model for Euler diagrams, a simple algorithm was used. Sophisticated force-directed algorithms, that could speed-up the layout process and effectively handle more criteria and more complex Euler diagrams, should now be adopted. Drawing methods that use a force-directed approach to automatically generate an Euler diagram with smooth curves and with only the required set of zones given any abstract description, should be devised. The proposed future work is discussed in detail in Section 4.5.

In Chapter 5, we have demonstrated the effectiveness of ellipses in drawing accurate, wellformed area-proportional 3-Venn diagrams with smooth curves for most random data. However, there are area specifications for which such a diagram cannot be drawn accurately using convex curves [Chow, 2007; Rodgers et al., 2009] and thus ellipses. So other shapes that are more general than ellipses (e.g., ovals) should be investigated, in that drawing methods that generate diagrams with such a curve shape are devised and the characteristics of the area specifications for which a diagram can be drawn accurately with that curve are identified. In this way, when an area-proportional diagram needs to be drawn for a given area specification, different shaped curves could be considered in a "natural progression" [Chow, 2007, p. 83], from the most specific to the most general (e.g., circles, ellipses, ovals, *n*-ellipses, regular

n-gons and irregular *n*-gons), until a curve shape that generates an accurate diagram for the required area specification is found. This would ensure that the simplest and smoothest curves that can accurately depict the required data are used, and that less desired and non-smooth curves such as polygons are among the last to be considered. If the characteristics of the area specifications for which an accurate diagram can be drawn using each of the shaped curves are identified, then given an area specification, the most appropriate curve shape would automatically be established without the need to try various shaped curves in the progression.

Being the first to use ellipses and considering the complexities in handling such shapes (e.g., to compute the zone areas and to adjust their properties) [Chow and Ruskey, 2004], we focused on area-proportional Venn diagrams with three curves. Our method eulerAPE already draws some Euler diagrams, but abstract descriptions and area specifications for which an accurate area-proportional 3-Euler diagram can be drawn with ellipses should be identified, and the effectiveness of eulerAPE in handling such cases should be evaluated. Further investigations are also required to identify whether accurate area-proportional Venn and Euler diagrams with more curves can be drawn accurately with ellipses. Further details on this and other future work are provided in Section 5.9.

In Chapter 6, our user study on the effect of different Euler diagram visualizations on Bayesian reasoning revealed no statistically significant differences between visualization types and thus, it is still unclear whether Euler diagrams with glyphs are more effective than area-proportional Euler diagrams. So further investigations should be carried out to identify effective ways how set relations and related quantitative data can be accurately and easily depicted.

The results of our study indicated that simply adding a diagram to the classic text of the Bayesian problem is not a viable solution. So other text+visualization presentation techniques should be considered and evaluated. For instance, the numbers and the text (story) could be shown on or around the visualization. Miniature visualizations like Tufte's [2006] sparklines could be embedded in the text, and storytelling techniques could be adopted. Such presentation techniques should also promote the use of estimation rather than calculation, as our results indicated a significant reduction in probability estimation errors when numerical values were replaced with phrases that just set the scene. After all, in such problems, a good estimation that is close to the exact calculated answer is adequate for making informed decisions. Further details are provided in Section 6.7.

7.3 Final Remark

Venn and Euler diagrams are amongst the oldest and most widely used visualizations for representing set-typed data. In contrast to other set visualizations, these diagrams use closed curves to depict relationships between the sets. Closed curves that have good continuity are preattentively processed, pop out and have a strong perceptual grouping effect that could aid reasoning about the sets and their relationships. However, it is still unclear when these diagrams are most helpful and how other aesthetic features of the diagrams could affect comprehension. Hence, empirical studies should be conducted to explore and evaluate the design space of these diagrams, provide guidance when to use these diagrams, and formalize aesthetic metrics of a 'good' Venn and Euler diagram that possibly "forces us to notice what we never expected to see" [Tukey, 1977].

Bibliography

James Abello and Frank van Ham (2004). Matrix zoom: A visual interface to semi-external graphs. In: *Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis)*, pp. 183-90, Austin, TX, USA. IEEE.

Agilent Technologies Inc (2012). *GeneSpring GX Software*. Accessed February 28, 2013, http://www.strandgenomics.com/GeneSpring.

Icek Ajzen (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, 35(5) : 303-14.

Md Jawaherul Alam, Therese Biedl, Stefan Felsner, Michael Kaufmann, Stephen G Kobourov, and Torsten Ueckerdt (2012). Computing Cartograms with Optimal Complexity. In: *Proceedings of the 28th Annual ACM Symposium on Computational Geometry*, pp. 21-30, Chapel Hill, NC, USA. ACM.

F K Aldrich and L Sheppard (2000). Graphicacy: the fourth 'R'? *Primary Science Review*, 64: 8-11.

Salvatore Alfano and Meredith L Greer (2003). Determining If Two Solid Ellipsoids Intersect. *Journal of Guidance, Control, and Dynamics*, 26(1) : 106-10.

Tony Allan (2011). *Virtual water: tackling the threat to our planet's most precious resource*, IB Tauris, London, UK (icons and illustrations by Angela Morelli).

Basak Alper, Nathalie Riche, Gonzalo Ramos, and Mary Czerwinski (2011). Design study of LineSets, a novel set visualization technique. *IEEE Transactions on Visualization and Computer Graphics*, 17(12) : 2259-67.

Bilal Alsallakh, Wolfgang Aigner, Silvia Miksch, and Helwig Hauser (2013). Radial Sets: Interactive Visual Analysis of Large Overlapping Sets. *IEEE Transactions on Visualization and Computer Graphics*, 19(12).

Jessica S Ancker, Yalini Senathirajah, Rita Kukafka, and Justin B Starren (2006). Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association*, 13(6) : 608-18.

Jessica S Ancker, Elke U Weber, and Rita Kukafka (2011). Effect of arrangement of stick figures on estimates of proportion in risk graphics. *Medical Decision Making*, 31(1) : 143-50.

H M Antia (2002). *Numerical methods for scientists and engineer, 2nd Edition*, Birkhäuser Verlag, Basel, Switzerland.

David Archer (2009). *The long thaw: how humans are changing the next 100,000 years of Earth's climate*, Princeton University Press, Princeton, NJ, USA.

Emma L Aronson, Eric A Dubinsky, and Brent R Helliker (2013). Effects of nitrogen addition on soil microbial diversity and methane cycling capacity depend on drainage conditions in a pine forest soil. *Soil Biology and Biochemistry*, 62: 119-28.

Franz Aurenhammer (1991). Voronoi diagrams — A Survey of a Fundamental Geometric Data Structure. *ACM Computing Surveys (CSUR)*, 23(3) : 345-405.

Stefan Axelsson (2000). The Base-Rate Fallacy and the Difficulty of Intrusion Detection. *ACM Transactions on Information and System Security (TISSEC)*, 3(3) : 186-205.

Michael J Axtell (2013). ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA*, 19(6) : 740-51.

Nathan W Bailey, Paris Veltsos, Yew-Foon Tan, A Harvey Millar, Michael G Ritchie, and Leigh W Simmons (2013). Tissue-Specific Transcriptomics in the Field Cricket Teleogryllus oceanicus. *G3: Genes, Genomes, Genetics*, 3(2) : 225-30.

John C Baird (1970). *Psychophysical Analysis of Visual Space*, Pergamon Press, New York, NY, USA.

Michael Balzer and Oliver Deussen (2007). Level-of-detail visualization of clustered graph layouts. In: *Proceedings of the 6th International Asia-Pacific Symposium on Visualization (APVIS)*, pp. 133-40, Sydney, Australia. IEEE.

Maya Bar-Hillel (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3) : 211-33.

Maya Bar-Hillel (1983). The base rate fallacy controversy. In Roland W Scholz (ed.), *Decision Making under Uncertainty,* Elsevier Science Publishers B V, Amsterdam, The Netherlands, pp. 39-61.

Inigo Barandiaran, Camilo Cortes, Marcos Nieto, Manuel Grana, and Oscar E Ruiz (2013a). A New Evaluation Framework and Image Dataset for Keypoint Extraction and Feature Descriptor Matching. In: *Proceedings of the 8th International Conference on Computer Vision Theory and Application (VISAPP)*, Barcelona, Spain.

Iñigo Barandiaran, Manuel Graña, and Marcos Nieto (2013b). An Empirical Evaluation of Interest Point Detectors. *Cybernetics and Systems: An International Journal*, 44(2-3) : 98-117.

Aron K Barbey and Steven A Sloman (2007a). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences*, 30(3) : 241-254, commentary 255.

Aron K Barbey and Steven A Sloman (2007b). Base-rate respect: From statistical formats to cognitive structures. *Behavioral and Brain Sciences*, 30(3) : 287-92.

Josh Barnes and Piet Hut (1986). A hierarchical O (N log N) force-calculation algorithm. *Nature*, 324: 446-49.

Margaret E Baron (1969). A note on the historical development of logic diagrams: Leibniz, Euler and Venn. *The Mathematical Gazette*, 53(384) : 113-25.

Daniel Bates and Hayley Peterson (2012), 'EXCLUSIVE: Gunman's 'survivalist' mother was OBSESSED with guns and stockpiled supplies as she planned for economic collapse before son shot her in the face while she lay in bed', *Daily Mail*, December 15, 2012.

Malcolm I Bauer and Philip Nicholas Johnson-Laird (1993). How Diagrams Can Improve Reasoning. *Psychological Science*, 4(6) : 372-78.

Thomas Bayes (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions of the Royal Society of London (1683-1775)*, 53: 370-418.

Chris Bennett, Jody Ryall, Leo Spalteholz, and Amy Gooch (2007). The Aesthetics of Graph Visualization. In: *Proceedings of the 3rd Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging (Computational Aesthetics)*, pp. 57-64, Aire-la-Ville, Switzerland. Eurographics Association.

Florence Benoy and Peter Rodgers (2007). Evaluating the comprehension of Euler diagrams. In: *Proceedings of the 11th International Conference on Information Visualization (IV)*, pp. 771-80, Zürich, Switzerland. IEEE Computer Society.

Brent Berlin and Paul Kay (1969). *Basic color terms: Their universality and evolution*, University of California Press, Berkeley, CA, USA (berlin1991basic).

Jacob Bernoulli (1713). *Ars conjectandi, opus posthumum: Accedit Tractatus de seriebus infinitis, et epistola Gallicé scripta de ludo pilae reticularis*, Thurnisii, Basel, Switzerland (English translation: Oscar Sheynin (2005), 'The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis', Johns Hopkins University Press, Baltimore, MD, USA).

Jacques Bertin (1983). *Semiology of Graphics*, University of Wisconsin Press, Madison, WI, USA.

Therese Biedl and Goos Kant (1998). A better heuristic for orthogonal graph drawings. *Computational Geometry*, 9(3) : 159-80.

Piotr Bielecki, Uliana Komor, Agata Bielecka, Mathias Müsken, Jacek Puchałka, Mathias W Pletz, Manfred Ballmann, dos Santos Martins, Vítor AP, Siegfried Weiss, and Susanne Häussler (2013). Ex vivo transcriptional profiling reveals a common set of genes important for the adaptation of Pseudomonas aeruginosa to chronically infected host sites. *Environmental Microbiology*, 15(2) : 570-87.

Alsallakh Bilal, Micallef Luana, Wolfgang Aigner, Helwig Hauser, Silvia Miksch, and Rodgers Peter (2014). Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. In: *Proceedings of the 16th Annual Eurographics Conference on Visualization (EuroVis), State of the Art Reports (STARs)*, Eurographics. http://www.setviz.net.

BioInfoRx Inc (2013). *Data Overlapping and Area-Proportional Venn Diagram*. BioInfoRx, Inc. Accessed May 21, 2013, http://bioinforx.com/lims/cloud-based-free-research-tools-for-scientific-data-management-and-analysis/bxtoolbox.

George D. Birkhoff (1933). *Aesthetic Measure*, Harvard University Press, Cambridge, MA, USA.

Andrew Blake, Gem Stapleton, Peter Rodgers, Liz Cheek, and John Howse (2012a). Does the Orientation of an Euler Diagram Affect User Comprehension? In: *Proceedings of the 18th International Conference on Distributed Multimedia Systems, International Workshop on Visual Languages and Computing (VLC)*, pp. 185-90, Miami Beach, FL, USA. Knowledge Systems Institute.

Andrew Blake, Gem Stapleton, Peter Rodgers, and John Howse (2012b). What Makes an Effective Euler Diagram? In: *Compendium of the 7th International Conference on Diagrammatic Representation and Inference (Diagrams), Graduate Student Symposium*, pp. 11-13, Canterbury, UK.

Randolph Blake and Karen Holopigian (1985). Orientation selectivity in cats and humans assessed by masking. *Vision Research*, 25(10) : 1459-67.

Matthew Bloch and Hannah Fairfield (2013), 'For the Elderly, Diseases That Overlap', *The New York Times*, April 15, 2013, sec. Science.

Wolfgang Böhm, Gerald Farin, and Jürgen Kahmann (1984). A survey of curve and surface methods in CAGD. *Computer Aided Geometric Design*, 1(1) : 1-60.

Sylvia F Boj, Joan Marc Servitja, David Martin, Martin Rios, Iannis Talianidis, Roderic Guigo, and Jorge Ferrer (2009). Functional targets of the monogenic diabetes transcription factors HNF-1α and HNF-4α are highly conserved between mice and humans. *Diabetes*, 58(5) : 1245-53.

John Adrian Bondy and Uppaluri Siva Ramachandra Murty (1976). *Graph Theory With Applications*, Macmillan Press, London, UK (bondy1976graph).

Ann Bostrom, Luc Anselin, and Jeremy Farris (2008). Visualizing Seismic Risk and Uncertainty. *Annals of the New York Academy of Sciences*, 1128(1) : 29-40.

Daniel Bouyer, Francois Roudier, Maren Heese, Ellen D Andersen, Delphine Gey, Moritz K Nowack, Justin Goodrich, Jean-Pierre Renou, Paul E Grini, Vincent Colot, and Arp Schnittger (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genetics*, 7(3) : e1002014.

William J. Bowman (1967). *Graphic Communication.*, Wiley Series on Human Communication, vol. 18, Wiley, New York, NY, USA.

Jürgen Branke, Frank Bucher, and Hartmut Schmeck (1997). Using genetic algorithms for drawing undirected graphs. In: *Proceedings of the 3rd Nordic Workshop on Genetic Algorithms and their Applications (3NWGA)*, Helsinki, Finland.

Gary L Brase (2002). Which statistical formats facilitate what decisions? The perception and influence of different statistical information formats. *Journal of Behavioral Decision Making*, 15(5) : 381-401.

Gary L Brase, Laurence Fiddick, and Clare Harries (2006). Participant recruitment methods and statistical reasoning performance. *The Quarterly Journal of Experimental Psychology*, 59(05) : 965-76.

Gary L Brase (2008). Frequency interpretation of ambiguous statistical information facilitates Bayesian reasoning. *Psychonomic Bulletin & Review*, 15(2) : 284-89.

Gary L Brase (2009). Pictorial Representations in Statistical Reasoning. *Applied Cognitive Psychology*, 23(3) : 369-81.

Richard Brath (2012). Multi-Attribute Glyphs on Venn and Euler Diagrams to Represent Data and Aid Visual Decoding. In: *Proceedings of the 3rd International Workshop on Euler Diagrams, CEUR-WS. org vol. 854*, pp. 122-29, Canterbury, UK.

Willard Cope Brinton (1916). *Graphic methods for presenting facts*, The Engineering Magazine Company, New York, NY, USA.

Sandra M Brown, Julie O Culver, Kathryn E Osann, Deborah J MacDonald, Sharon Sand, Andrea A Thornton, Marcia Grant, Deborah J Bowen, Kelly A Metcalfe, Harry B Burke, Mark E Robson, Susan Friedman, and Jeffrey N Weitzel (2011). Health literacy, numeracy, and interpretation of graphical breast cancer risk estimates. *Patient Education and Counseling*, 83: 92-98.

Mark Bruls, Kees Huizing, and Jarke J van Wijk (2000). Squarified Treemaps. In: *Proceedings of Joint Eurographics and IEEE TCVG Symposium on Visualization (TCVG 2000)*, pp. 33-42, Springer.

Mark Bruls, Johan Geerlings, Frank van Ham, Kees Huizing, Elisabeth Melby, Huub van de Wetering, and Jarke J van Wijk (2002). *SequoiaView*. Eindhoven University of Technology, Eindhoven, Netherlands. Accessed August 31, 2013, http://w3.win.tue.nl/nl/onderzoek/onderzoek_informatica/visualization/sequoiaview.

Kevin Buchin, Bettina Speckmann, and Sander Verdonschot (2012). Evolution Strategies for Optimizing Rectangular Cartograms. *Geographic Information Science, Lecture Notes in Computer Science*, 7478: 29-42.

Ed Bullmore and Olaf Sporns (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3) : 186-98.

Richard L Burden and John Douglas Faires (2010). Solutions of Equations in One Variable, The Bisection Method. *Numerical Analysis, 9th Edition,* Cengage Learning, Boston, MA, USA, pp. 48-54.

Francis John Burdon, Angus R McIntosh, and Jon S Harding (2013). Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. *Ecological Applications*, 23: 1036-47.

Kevin Burns (2003). Improving Intuition with Bayesian Boxes. In: *Proceedings of the 6th International Conference on Naturalistic Decision Making (NDM)*, Pensacola Beach, FL, USA.

Kevin Burns (2004a). Painting pictures to augment advice. In: *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pp. 344-49, Gallipoli, Italy. ACM.

Kevin Burns (2004b). Bayesian Boxes: a colored calculator for picturing posteriors. In: *Proceedings of the 3rd International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 2980*, pp. 382-84, Cambridge, UK. Springer.

Oliver Byrne (1847). *The First Six Books of the Elements of Euclid in which coloured diagrams and symbols are used instead of letters for the greater ease of learners*, William Pickering, London, UK.

Huan Cai, Hongyu Chen, Tie Yi, Caitlin M Daimon, John P Boyle, Chris Peers, Stuart Maudsley, and Bronwen Martin (2013). VennPlex—A Novel Venn Diagram Program for Comparing and Visualizing Datasets with Differentially Regulated Datapoints. *PLoS ONE*, 8(1) : e53388.

Cancer Research UK (2012). *Breast Screening Statistics*, CancerStats: Cancer Statistics for the UK. Cancer Research UK, London, UK. Accessed March 23, 2013, http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/screening.

Stuart K Card, George G Robertson, and Jock D Mackinlay (1991). The information visualizer, an information workspace. In: *Proceedings of the 9th SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 181-86, New Orleans, LA, USA. ACM.

A J Carr (1996). A patient-centred approach to evaluation and treatment in rheumatoid arthritis: the development of a clinical tool to measure patient-perceived handicap. *Rheumatology*, 35(10) : 921-32.

Jeremy Carroll, Frank Ruskey, and Mark Weston (2007). Which n-Venn diagrams can be drawn with convex k-gons? *Discrete & Computational Geometry*, 37(4) : 619-28.

Jeremy J Carroll (2000), *Drawing Venn Triangles*, Technical Report HPL-2000-73, Hewlett-Packard Laboratories, Bristol, UK.

Paulo C Carvalho, Juliana S G Fischer, Emily I Chen, John R Yates, and Valmir C Barbosa (2008). PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics*, 9(1) : 316.

Ward Casscells, Arno Schoenberger, and Thomas B Graboys (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine*, 299(18) : 999-1001.

Nick Cawthon and A Vande Moere (2006). A conceptual model for evaluating aesthetic effect within the user experience of information visualization. In: *Proceedings of the 10th International Conference on Information Visualization (IV)*, pp. 374-82, London, UK. IEEE Computer Society.

Nick Cawthon and Andrew Vande Moere (2007). The Effect of Aesthetic on the Usability of Data Visualization. In: *Proceedings of the 11th International Conference on Information Visualization (IV)*, pp. 637-48, Zurich, Switzerland. IEEE Computer Society.

CENSUS (2007). *Census of Antique Works of Art and Architecture Known in the Renaissance*. Berlin-Brandenburgische Akademie der Wissenschaften and Humboldt-Universität zu Berlin. Accessed August 31, 2013, www.census.de.

Paul Chandler and John Sweller (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4) : 293-332.

Gretchen B Chapman and Jingjing Liu (2009). Numeracy, frequency, and Bayesian reasoning. *Judgment and Decision Making*, 4(1) : 34-40.

Nick Chater (1997). Simplicity and the mind. *The Psychologist*, 10(11) : 495-98.

Hanbo Chen and Paul C Boutros (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, 12(1) : 35.

Jenny J Chen, Natala J Menezes, Adam D Bradley, and T A North (2011). Opportunities for Crowdsourcing Research on Amazon Mechanical Turk. *Human Factors*, 5: 3.

Peter C-H Cheng and Nigel G Pitt (2003). Diagrams for difficult problems in probability. *Mathematical Gazette*, 87(508) : 86-97.

Chicago Tribune (1993), 'A false hiv test caused 18 months of hell', *Chicago Tribune*, March 5, 1993.

Stirling Chow and Frank Ruskey (2004). Drawing Area-Proportional Venn and Euler Diagrams. In: *Proceedings of the 11th International Symposium on Graph Drawing (GD 2003), Lecture Notes in Computer Science 2912*, pp. 466-77, Perugia, Italy. Springer.

Stirling Chow and Peter Rodgers (2005). Constructing area-proportional Venn and Euler diagrams with three circles. In: *Proceedings of the 2nd International Workshop on Euler Diagrams*, Paris, France.

Stirling Chow and Frank Ruskey (2005). Towards a general solution to drawing area-proportional Euler diagrams. *Electronic Notes in Theoretical Computer Science*, 134: 3-18.

Stirling Chow and Frank Ruskey (2007). Minimum area Venn diagrams whose curves are polyominoes. *Mathematics Magazine*, 80(2) : 91-103.

Stirling Christopher Chow (2007), *Generating and Drawing Area-Proportional Venn and Euler Diagrams*, PhD thesis, Department of Computer Science, University of Victoria, Victoria, BC, Canada.

Mohcine Chraibi, Armin Seyfried, and Andreas Schadschneider (2010). Generalized centrifugal-force model for pedestrian dynamics. *Physical Review E*, 82(4) : 046111.

Mohcine Chraibi (2012). *Validated force-based modeling of pedestrian dynamics*, University of Cologne, Cologne, Germany.

Hyun Bae Chun, Michael Scott, Sherry Niessen, Heather Hoover, Andrew Baird, John Yates, Bruce E Torbett, and Brian P Eliceiri (2011). The proteome of mouse brain microvessel membranes and basal lamina. *Journal of Cerebral Blood Flow & Metabolism*, 31(12) : 2267-81.

Andrew K Clark (1976). Re-evaluation of Monte Carlo studies in nonmetric multidimensional scaling. *Psychometrika*, 41(3) : 401-3.

Jeff Clark (2008). *Twitter Venn*, Neoformix. Accessed February 28, 2013, http://www.neoformix.com/2008/TwitterVenn.html.

Michael J Clark, Rui Chen, Hugo YK Lam, Konrad J Karczewski, Rong Chen, Ghia Euskirchen, Atul J Butte, and Michael Snyder (2011). Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology*, 29(10) : 908-14.

William S Cleveland, Charles S Harris, and Robert McGill (1982). Judgments of circle sizes on statistical maps. *Journal of the American Statistical Association*, 77(379) : 541-47.

William S Cleveland and Robert McGill (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387) : 531-54.

William S Cleveland (1985). *The Elements of Graphing Data, 1st Edition*, Wadsworth Advanced Books and Software, Monterey, California.

William S Cleveland and Robert McGill (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716) : 828-33.

William S Cleveland and Robert McGill (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society*, Series A(General) : 192-229.

William S Cleveland (1994). *The Elements of Graphing Data, Revised Edition*, Hobart Press, Summit, NJ, USA.

Jack Cohen (1988). *Statistical power analysis for the behavioral sciencies, 2nd Edition*, Lawrence Erlbaum Associates, New York, NY, USA.

Jacob Cohen (1992). A Power Primer. *Psychological Bulletin*, 112(1) : 155.

Jacob Cohen (1994). The earth is round (p<. 05). *American Psychologist*, 49(12) : 997-1003.

William G Cole (1989). Understanding Bayesian Reasoning Via Graphical Displays. In: *Proceedings of the 7th SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 381-86, Austin, TX, USA. ACM.

William G Cole and Janet E Davidson (1989). Graphic Representation Can Lead to Fast and Accurate Bayesian Reasoning. In: *Proceedings of the 13th Annual Symposium on Computer Application in Medical Care*, pp. 227-31, Washington, DC, USA.

Michael K Coleman and D Stott Parker (1996). Aesthetics-based Graph Layout for Human Consumption. *Software: Practice and Experience*, 26(12) : 1415-38.

Christopher Collins, Gerald Penn, and Sheelagh Carpendale (2009). Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6) : 1009-16.

Michelle Patrick Cook (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, 90(6) : 1073-91.

Gennaro Cordasco, Rosario De Chiara, Andrew Fish, and Vittorio Scarano (2012). FunEuler: an Euler Diagram based Interface Enhanced with Region-based Functionalities. In: *Proceedings of the 3rd International Workshop on Euler Diagrams, CEUR-WS. org vol. 854*, pp. 107-21, Canterbury, UK.

Audrey Coreau and Jean-Louis Martin (2007). Multi-scale study of bird species distribution and of their response to vegetation change: a Mediterranean example. *Landscape Ecology*, 22(5) : 747-64.

Stanley Coren and Joan Stern Girgus (1978). *Seeing is deceiving: The psychology of visual illusions*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

Leda Cosmides and John Tooby (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58(1) : 1-73.

Louis Couturat (1901). *La logique de Leibniz*, Paris, France.

Louis Couturat (1903). *Opuscules et fragments inédits de Leibniz*, Paris, France.

Nelson Cowan (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1) : 87-185.

Nelson Cowan, Emily M Elliott, Saults Scott, J, Candice C Morey, Sam Mattox, Anna Hismjatullina, and Andrew RA Conway (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51(1) : 42-100.

Dena S Cox, Anthony D Cox, Lynne Sturm, and Greg Zimet (2010). Behavioral Interventions to Increase HPV Vaccination Acceptability Among Mothers of Young Girls. *Health Psychology*, 29(1) : 29-39.

Richard Cox (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9(4) : 343-63.

Paul V Crawford (1973). The Perception of Graduated Squares as Cartographic Symbols. *The Cartographic Journal*, 10(2) : 85-88.

Matthew A Cronin, Cleotilde Gonzalez, and John D Sterman (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes*, 108(1) : 116-30.

Kenneth Cukier (2010), 'Show me: New ways of visualising data', *The Economist*, February 27, 2010.

Angélique D'Hont, France Denoeud, Jean-Marc Aury, Franc-Christophe Baurens, Françoise Carreel, Olivier Garsmeur, Benjamin Noel, Stéphanie Bocs, Gaëtan Droc, Mathieu Rouard, Corinne Da Silva, Kamel Jabbari, Céline Cardi, Julie Poulain, Marlène Souquet, Karine Labadie, Cyril Jourda, Juliette Lengellé, Marguerite Rodier-Goud, Adriana Alberti, Maria Bernard, Margot Correa, Saravanaraj Ayyampalayam, Michael R Mckain, Jim Leebens-Mack, Diane Burgess, Mike Freeling, Didier Mbéguié-A-Mbéguié, Matthieu Chabannes, Thomas Wicker, Olivier Panaud, Jose Barbosa, Eva Hribova, Pat Heslop-Harrison, Rémy Habas, Ronan Rivallan, Philippe Francois, Claire Poiron, Andrzej Kilian, Dheema Burthia, Christophe Jenny, Frédéric Bakry, Spencer Brown, Valentin Guignon, Gert Kema, Miguel Dita, Cees Waalwijk, Steeve Joseph, Anne Dievart, Olivier Jaillon, Julie Leclercq, Xavier Argout, Eric Lyons, Ana Almeida, Mouna Jeridi, Jaroslav Dolezel, Nicolas Roux, Ange-Marie Risterucci, Jean Weissenbach, Manuel Ruiz, Jean-Christophe Glaszmann, Francis Quétier, Nabila Yahiaoui, and Patrick Wincker (2012). The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature*, 488(7410) : 213-17.

Leonardo da Vinci (1956). *Treatise on painting (Codex Urbinas Latinus 1270), vol. 2, facsimile*, Princeton University Press, Princeton, NJ, USA.

Tuan Dang, Anushka Anand, and Leland Wilkinson (2012). FmFinder: Search and Filter Your Favorite Songs. *Advances in Visual Computing, Lecture Notes in Computer Science*, 7431: 348-58.

Darik (2011). *VENN*, Matlab Central. The MathWorks Inc. Accessed February 28, 2013, http://www.mathworks.com/matlabcentral/fileexchange/22282-venn.

Ron Davidson and David Harel (1996). Drawing graphs nicely using simulated annealing. *ACM Transactions on Graphics*, 15(4) : 301-31.

Terry C Davis, Nancy C Dolan, M Rosario Ferreira, Cecilia Tomori, Kristen W Green, Alison M Sipler, and Charles L Bennett (2001). The role of inadequate health literacy skills in colorectal cancer screening. *Cancer Investigation*, 19(2) : 193-200.

Mark de Berg, Elena Mumford, and Bettina Speckmann (2010). Optimal BSPs and rectilinear cartograms. *International Journal of Computational Geometry & Applications*, 20(2) : 203-22.

Rosario De Chiara, Ugo Erra, and Vittorio Scarano (2003). VENNFS: A Venn-diagram file manager. In: *Proceedings of the 7th International Conference on Information Visualization (IV)*, pp. 120-25, London, UK. IEEE Computer Society.

Robin-Jan de Lange (2011). *Segregation-of-duties analysis apparatus and method*. US 7,941,336 B1, May 10. D2C Solutions, LLC, USA.

Aidan Delaney, Beryl Plimmer, Gem Stapleton, and Peter Rodgers (2010). Recognizing sketches of Euler diagrams drawn with ellipses. In: *Proceedings of the 16th International Conference on Distributed Multimedia Systems (DMS)*, pp. 305-10, Oak Brook, IL, USA.

Veronika Denes-Raj, Seymour Epstein, and Jonathan Cole (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 21(10) : 1083-92.

Robert F Dennehy (2001). The Springboard: How Storytelling Ignites Action in Knowledge-era Organizations. *Journal of Organizational Change Management*, 14(6) : 609-14.

René Descartes (1644). *Principia Philosophiae*, Apud Ludovicum Elzevirium, Amsterdam, The Netherlands.

Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G Tollis (1994). Algorithms for Drawing Graphs: an Annotated Bibliography. *Computational Geometry*, 4(5) : 235-82.

Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G Tollis (1999a). Force-Directed Methods. *Graph drawing: algorithms for the visualization of graphs,* Prentice-Hall, Upper Saddle River, NJ, USA, pp. 303-25.

Giuseppe Di Battista, Peter Eades, Roberto Tamassia, and Ioannis G Tollis (1999b). *Graph drawing: algorithms for the visualization of graphs*, Prentice-Hall, Upper Saddle River, NJ, USA.

Josep Díaz, Jordi Petit, and Maria Serna (2002). A Survey of Graph Layout Problems. *ACM Computing Surveys (CSUR)*, 34(3) : 313-56.

Kasper Dinkla, Marc J van Kreveld, Bettina Speckmann, and Michel A Westenberg (2012). Kelp Diagrams: Point Set Membership Visualization. *Computer Graphics Forum*, 31(3pt1) : 875-84.

James G Dolan and Stephen Iadarola (2008). Risk communication formats for low probability events: an exploratory study of patient preferences. *BMC Medical Informatics and Decision Making*, 8: 14.

Carol M Donnelly and Mark A McDaniel (1993). Use of analogy in learning scientific concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4) : 975-87.

Daniel Dorling (1996). *Area Cartograms: Their Use and Creation*, Concepts and Techniques in Modern Geography (CATMOG), vol. 59, School of Environmental Sciences, University of East Anglia, Norwich, UK.

James A Chrisman Dougenik, Nicholas R Niemeyer, Duane R (1985). An algorithm to construct continuous area cartograms. *The Professional Geographer*, 37(1) : 75-81.

Geoffrey M Draper, Yarden Livnat, and Richard F Riesenfeld (2009). A Survey of Radial Methods for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(5) : 759-76.

William Dunham (1999). *Euler: The master of us all*, Dolciani Mathematical Expositions No. 22, The Mathematical Association of America, USA.

Cody Dunne and Ben Shneiderman (2009), *Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts*, Technical Report HCIL-2009-13, University of Maryland, College Park, MD, USA.

Clement Vavasor Durell (1936). *Elementary Geometry*, G. Bell and Sons, London, UK.

Tim Dwyer (2001). Three dimensional UML using force directed layout. In: *Proceedings of the 1st Asia-Pacific Symposium on Information Visualisation (APVIS), CRPIT vol. 9*, pp. 77-85, Syndey, Australia. Australian Computer Society.

Tim Dwyer, Bongshin Lee, Danyel Fisher, Kori Inkpen Quinn, Petra Isenberg, George Robertson, and Chris North (2009). A Comparison of User-Generated and Automatic Graph Layouts. *IEEE Transactions on Visualization and Computer Graphics*, 15(6) : 961-68.

Carmel Bitondo Dyer, James S Goodwin, Sabrina Pickens-Pace, Jason Burnett, and P Adam Kelly (2007). Self-neglect among the elderly: a model based on more than 500 patients seen by a geriatric medicine team. *American Journal of Public Health*, 97(9) : 1671-76.

Arnaud Dzeing-Ella, Pascal C Nze Obiang, Rose Tchoua, Timothy Planche, Béatrice Mboza, Monique Mbounja, Ulrich Muller-Roemer, Joseph Jarvis, Eric Kendjo, Edouard Ngou-Milama, Peter G Kremsner, Sanjeev Krishna, and Maryvonne Kombila (2005). Severe falciparum malaria in Gabonese children: clinical and laboratory features. *Malaria Journal*, 4(1).

Peter Eades (1984). A Heuristic for Graph Drawing. *Congressus Numerantium*, 42: 149-60.

Peter Eades, Wei Lai, Kazuo Misue, and Kozo Sugiyama (1991), *Preserving the Mental Map of a Diagram*, Technical Report IIAS-RR-91-16E, International Institute for Advanced Study of Social Information Science, Fujitsu Limited, Numazu-shi, Shizuoka, Japan.

Peter Eades, Xuemin Lin, and Roberto Tamassia (1996). An algorithm for drawing a hierarchical graph. *International Journal of Computational Geometry and Applications*, 6(2) : 145-56.

Peter Eades, Qing-Wen Feng, and Xuemin Lin (1997). Straight-line drawing algorithms for hierarchical graphs and clustered graphs

Graph drawing.: 113-28.

Peter Eades and Xuemin Lin (2000). Spring Algorithms and Symmetry. *Theoretical Computer Science*, 240(2) : 379-405.

David Eberly (2010). *The Area of Intersecting Ellipses*. Geometric Tools, LLC. Accessed July 10, 2013, http://www.geometrictools.com/Documentation/AreaIntersectingEllipses.pdf.

David Eberly (2011a). *Information About Ellipses*. Geometric Tools, LLC. Accessed July 10, 2013, http://www.geometrictools.com/Documentation/InformationAboutEllipses.pdf.

David Eberly (2011b). *Intersection of Ellipses*. Geometric Tools, LLC. Accessed July 10, 2013, http://www.geometrictools.com/Documentation/IntersectionOfEllipses.pdf.

Daniel M Eddy (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.), *Judgment under uncertainty: Heuristics and biases,* Cambridge University Press, Cambridge, UK, pp. 249-67.

Anthony W F Edwards (1989). Venn diagrams for many sets. *New Scientist*, 7: 51-56.

Hartmut Ehrig, Gregor Engels, Hans-Jörg Kreowski, and Grzegorz Rozenberg (1999). *Handbook of Graph Grammars and Computing by Graph Transformation: Applications, Languages and Tools*, vol. 2, World Scientific Publishing Co, River Edge, NJ, USA.

Ruth B Ekstrom, John W French, Harry Horace Harman, and David Dermen (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*, Educational Testing Service, Princeton, NJ, USA.

Hamid el Azzouzi, Stefanos Leptidis, Meriem Bourajjaj, Anne-Sophie Armand, Roel van der Nagel, Marc van Bilsen, Paula A Da Costa Martins, and Leon J De Windt (2011). Peroxisome proliferator-activated receptor (PPAR) gene profiling uncovers insulin-like growth factor-1 as a PPARα target gene in cardioprotection. *Journal of Biological Chemistry*, 286(16) : 14598-607.

Joann G Elmore and Gerd Gigerenzer (2005). Benign Breast Disease—The Risks of Communicating Risk. *The New England Journal of Medicine*, 353(3) : 297-99.

Linda S Elting, Charles G Martin, Scott B Cantor, and Edward B Rubenstein (1999). Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ*, 318(7197) : 1527-31.

Randall W Engle, Michael J Kane, and Stephen W Tuholski (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In Akria Miyake and Priti Shah (eds.), *Models of working memory: Mechanisms of active maintenaance and executive control,* Cambridge University Press, Cambridge, UK, pp. 102-34.

James Enns (1986). Seeing textons in context. *Perception & Psychophysics*, 39(2) : 143-47.

Carlos Estrada, Vetta Barnes, Cathy Collins, and James C Byrd (1999). Health literacy and numeracy. *JAMA: The Journal of the American Medical Association*, 282(6) : 527-527.

Leonhard Euler (1768a). *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie, vol. 2, Lettres 102-108*, L'Académie Impériale des Sciences de Saint-Pétersbourg, St Petersburg, Russia.

Leonhard Euler (1768b). *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie*, L'Académie Impériale des Sciences de Saint-Pétersbourg, St Petersburg, Russia (1768-1772).

Leonhard Euler (1795). *Letters to a German Princess, on different subjects in physics and philosophy*, Translated from French by Henry Hunter, London, UK (Original: 'Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie', 1768-1772, L'Académie Impériale des Sciences de Saint-Pétersbourg, St Petersburg, Russia).

Jonathan St B T Evans, Simon J Handley, Nick Perham, David E Over, and Valerie A Thompson (2000). Frequency versus probability formats in statistical word problems. *Cognition*, 77(3) : 197-213.

Angela Fagerlin, Catharine Wang, and Peter A Ubel (2005). Reducing the Influence of Anecdotal Reasoning on People's Health Care Decisions: is a Picture Worth a Thousand Statistics? *Medical Decision Making*, 25(4) : 398-405.

Angela Fagerlin, Peter A Ubel, Dylan M Smith, and Brian J Zikmund-Fisher (2007a). Making numbers matter: Present and future research in risk communication. *American Journal of Health Behavior*, 31(Supplement 1) : S47-56.

Angela Fagerlin, Brian J Zikmund-Fisher, Peter A Ubel, Aleksandra Jankovic, Holly A Derry, and Dylan M Smith (2007b). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, 27(5) : 672-80.

Ruma Falk (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition*, 43(3) : 197-223.

Hong Fang, Stephen C Harris, Zhenjiang Su, Minjun Chen, Feng Qian, Leming Shi, Roger Perkins, and Weida Tong (2009). ArrayTrack: an FDA and public genomic tool. *Methods in Molecular Biology*, 563: 379-98.

Gerald E Farin, Josef Hoschek, and Myung-Soo Kim (2002). *Handbook of Computer Aided Geometric Design*, Elsevier (North-Holland), Amsterdam, The Netherlands.

Graham Farrell and William Sousa (2001). Repeat victimization and hot spots: the overlap and its implications for crime control and problem-orientated policing. *Crime Prevention Studies*, 12: 221-40.

Andy Field (2009). *Discovering Statistics Using SPSS, 3rd Edition*, Sage Publications, London, UK.

David J Field, Anthony Hayes, and Robert F Hess (1993). Contour Integration by the Human Visual System: Evidence for a Local "Association Field". *Vision Research*, 33(2) : 173-93.

Daniel Filonik and Dominikus Baur (2009). Measuring Aesthetics for Information Visualization. In: *Proceedings of the 13th International Conference on Information Visualization (IV)*, pp. 579 -584, Barcelona, Spain. IEEE Computer Society.

Andrew Fish (2009). Euler Diagram Transformations. *Proceedings of the 8th International Workshop on Graph Transformation and Visual Modeling Techniques (GT-VMT), Electronic Communications of the EASST vol. 18.*

Andrew Fish, Babak Khazaei, and Chris Roast (2011). User-comprehension of Euler diagrams. *Journal of Visual Languages & Computing*, 22(5) : 340-54.

Clifford J Fisk, David L Caskey, and Leslie E West (1967). ACCEL: Automated circuit card etching layout. *Proceedings of the IEEE*, 55(11) : 1971-82.

Paul Fitzpatrick (1960). Leading British statisticians of the nineteenth century. *Journal of the American Statistical Association*, 55(289) : 38-70.

James John Flannery (1971). The relative effectiveness of some common graduated point symbols in the presentation of quantitative point data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 8(2) : 96-109.

Jean Flower and John Howse (2002). Generating Euler Diagrams. In: *Proceedings of the 2nd International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 2317*, pp. 61-75, Callaway Gardens, GA, USA. Springer.

Jean Flower, John Howse, and John Taylor (2003a). Nesting in Euler diagrams. In: *Proceedings of the 1st International Workshop on Graph Transformation and Visual Modeling Techniques (GT-VMT 2002), Electronic Notes in Theoretical Computer Science vol. 72 no. 3*, pp. 93-102, Barcelona, Spain. Elsevier.

Jean Flower, Peter Rodgers, and Paul Mutton (2003b). Layout Metrics for Euler Diagrams. In: *Proceedings of the 7th International Conference on Information Visualization (IV)*, pp. 272-80, London, UK. IEEE Computer Society.

Jean Flower, Andrew Fish, and John Howse (2008). Euler Diagram Generation. *Journal of Visual Languages & Computing*, 19(6) : 675-94.

Heather L Flowers, Frank L Silver, Jiming Fang, Elizabeth Rochon, and Rosemary Martino (2013). The Incidence, Co-occurrence, and Predictors of Dysphagia, Dysarthria, and Aphasia after First-Ever Acute Ischemic Stroke. *Journal of Communication Disorders*, in press.

Charles Forceville (2002). *Pictorial metaphor in advertising*, Routledge, London, UK.

Dimitris Fouskakis and David Draper (2002). Stochastic optimization: a review. *International Statistical Review*, 70(3) : 315-49.

Michael Friendly and Daniel J Denis (2001). Milestones in the history of thematic cartography, statistical graphics, and data visualization. In: *Web document, accessed: November 19, 2012*, http://www.datavis.ca/milestones.

Michael Friendly (2008). A Brief History of Data Visualization. *Handbook of Data Visualization, Springer Handbooks Computational Statistics*, Springer, Heidelberg, Germany, pp. 15-56.

Thomas M J Fruchterman and Edward M Reingold (1991). Graph Drawing by Force-directed Placement. *Software: Practice and Experience*, 21(11) : 1129-64.

Richard Fuller, Nigel Dudley, and Jon Blacktop (2001). Risk communication and older people—understanding of probability and risk information by medical inpatients aged 75 years and older. *Age and Ageing*, 30(6) : 473-76.

Richard Fuller, Nigel Dudley, and Jon Blacktop (2002). How informed is consent? Understanding of pictorial and verbal probability information by medical inpatients. *Postgraduate Medical Journal*, 78(923) : 543-44.

Wolfgang Gaissmaier, Odette Wegwarth, David Skopec, Ann-Sophie Müller, Sebastian Broschinski, and Mary C Politi (2012). Numbers can be worth a thousand pictures: Individual differences in understanding graphical and numerical representations of health-related information. *Health Psychology*, 31(3) : 286-96.

Pawel Gajer, Michael T Goodrich, and Stephen G Kobourov (2001). A multi-dimensional approach to force-directed layouts of large graphs. In: *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000), Lecture Notes in Computer Science 1984*, pp. 211-21, Colonial Williamsburg, VA, USA. Springer.

Mirta Galesic, Rocio Garcia-Retamero, and Gerd Gigerenzer (2009). Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*, 28(2) : 210-16.

Mirta Galesic and Rocio Garcia-Retamero (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, 170(5) : 462-68.

Mirta Galesic and Rocio Garcia-Retamero (2011). Graph Literacy A Cross-Cultural Comparison. *Medical Decision Making*, 31(3) : 444-57.

Galileo Galilei (1613). *Istoria e dimostrazioni intorno alle macchie solari e loro accidenti... Si aggiungono nel fine le Lettere, e Disquisizioni del finto Apelle.*, Appresso Giacomo Mascardi, Rome, Italy.

Emden Gansner, Yifan Hu, Stephen Kobourov, and Chris Volinsky (2009). Putting recommendations on the map: visualizing clusters and relations. In: *Proceedings of the 3rd ACM Conference on Recommender Systems*, pp. 345-48, New York, NY, USA. ACM.

Emden R Gansner, Yifan Hu, and Stephen Kobourov (2010a). GMap: Visualizing graphs and clusters as maps. In: *Proceedings of the 3rd IEEE Pacific Visualization Symposium (PacificVis)*, pp. 201-8, Taipei, Taiwan. IEEE.

Emden R Gansner, Yifan Hu, and Stephen G Kobourov (2010b). Visualizing graphs and clusters as maps. *IEEE Computer Graphics and Applications*, 30(6) : 54-66.

Gapminder (2013). *Gapminder: Unveiling the beauty of statistics for a fact based world view*. Accessed March 30, 2013, http://www.gapminder.org.

Rocio Garcia-Retamero, Mirta Galesic, and Gerd Gigerenzer (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making*, 30(6) : 672-84.

Rocio Garcia-Retamero and Edward T Cokely (2011). Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied*, 17(3) : 270-87.

Rocio Garcia-Retamero and Mandeep K Dhami (2011). Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. *Health Expectations*, 14(S1) : 46-57.

Rocio Garcia-Retamero and Mirta Galesic (2013). *Transparent Communication of Health Risk: Overcoming Cultural Differences*, Springer, New York, NY, USA.

Rocio Garcia-Retamero and Ulrich Hoffrage (2013). Visual Representation of Statistical Information Improves Diagnostic Inferences in Doctors and Their Patients. *Social Science & Medicine*.

Mario R Garcia and Margaret Mary Stark (1991). *Eyes on the News*, Poynter Institute for Media Studies, St Petersburg, FL, USA.

Ken Garland (1994). *Mr Beck's Underground Map: a history*, Capital Transport Publishing, Harrow Weald, UK.

Ian P Gent and Toby Walsh (1993). Towards an understanding of hill-climbing procedures for SAT. *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI)*, 93: 28-33.

Nahum Gershon and Stephen G Eick (1995). Visualization's new tack: Making sense of information. *Spectrum, IEEE*, 32(11) : 38-40.

Nahum Gershon and Ward Page (2001). What storytelling can do for information visualization. *Communications of the ACM*, 44(8) : 31-37.

Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola (2004). A comparison of the readability of graphs using node-link and matrix-based representations. In: *Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis)*, pp. 17-24, Austin, TX, USA. IEEE.

Helen Gibson, Joe Faith, and Paul Vickers (2012). A survey of two-dimensional graph layout techniques for information visualisation. *Information Visualization*.

Gerd Gigerenzer, Wolfgang Hell, and Hartmut Blank (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3) : 513-25.

Gerd Gigerenzer and Ulrich Hoffrage (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review*, 102(4) : 684-704.

Gerd Gigerenzer (1998). Ecological Intelligence: An Adaptation for Frequencies. In Denise Dellarosa Cummins and Colin Allen (eds.), *The Evolution of Mind,* Oxford University Press, New York, NY, USA, pp. 9-29.

Gerd Gigerenzer and Ulrich Hoffrage (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review*, 106(2) : 425-30.

Gerd Gigerenzer, Peter M Todd, and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*, Oxford University Press, New York, NY, USA.

Gerd Gigerenzer and Reinhard Selten (2002). *Bounded rationality: The adaptive toolbox*, Mit Press, Cambridge, MA, USA.

Gerd Gigerenzer (2007). *Gut feelings: The intelligence of the unconscious*, Viking, New York, NY, USA.

Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M Schwartz, and Steven Woloshin (2007). Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, 8(2) : 53-96.

Gerd Gigerenzer (2010). Collective Statistical Illiteracy: A Cross-Cultural Comparison With Probabilistic National Samples: Comment on "Statistical Numeracy for Health". *Archives of Internal Medicine*, 170(5) : 468.

Vittorio Girotto and Michel Gonzalez (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition*, 78(3) : 247-76.

Vittorio Girotto and Michel Gonzalez (2002). Chances and frequencies in probabilistic reasoning: rejoinder to Hoffrage, Gigerenzer, Krauss, and Martignon. *Cognition*, 84(3) : 353-59.

Andrew S Glassner (2004). *Interactive Storytelling: Techniques for 21st Century Fiction*, A K Peters, Coyote Wind, LLC, Canada.

Fern Gobet and Gary Clarkson (2004). Chunks in expert memory: Evidence for the magical number four… or is it two? *Memory*, 12(6) : 732-47.

Wenfeng Gong and Jan Ostermann (2011). *PVENN: Stata module to create proportional Venn diagram*, Statistical Software Components S457368. Boston College Department of Economics, Chestnut Hill, MA, USA. Accessed February 28, 2013, http://ideas.repec.org/c/boc/bocode/s457368.html.

Irving John Good (1995). When batterer turns murderer. *Nature*, 375(6532) : 541.

Felicity Goodyear-Smith, Bruce Arroll, Lydia Chan, Rod Jackson, Sue Wells, and Timothy Kenealy (2008). Patients prefer pictures to numbers to express cardiovascular benefit from treatment. *The Annals of Family Medicine*, 6(3) : 213-17.

Google Inc (2012). *Venn Charts*, Google Chart Tool: Image Charts (deprecated). Accessed February 28, 2013, https://developers.google.com/chart/image/docs/gallery/venn_charts.

Timothy E Gookin, Junhyong Kim, and Sarah M Assmann (2008). Whole proteome identification of plant candidate G-protein coupled receptors in Arabidopsis, rice, and poplar: computational prediction and in-vivo protein coupling. *Genome Biology*, 9(7) : R120.

John D Gould (1976). Looking at Pictures. *Eye Movements and Psychological Processes*, : 323-45.

Richard L Gregory (1966). Eye and brain: The psychology of seeing, 5th Edition.

Mark S Gresnigt, Katharina L Becker, Sanne P Smeekens, Cor W M Jacobs, Leo A B Joosten, der Meer van, Jos W M, Mihai G Netea, and Frank L van de Veerdonk (2013). *Aspergillus fumigatus*—Induced IL-22 Is Not Restricted to a Specific Th Cell Subset and Is Dependent on Complement Receptor 3. *The Journal of Immunology*, 190(11) : 5629-39.

Jeffrey L Griffin and Robert L Stevenson (1992). Influence of Text and Graphics in Increasing Understanding of Foreign News Context. *Newspaper Research Journal*, 13(1) : 84-99.

Jeffrey L Griffin and Robert L Stevenson (1994a). The Effectiveness of "How Graphics" and Text in Presenting the News. *Visual Communication Quarterly*, 1(2) : 10-16.

Jeffrey L Griffin and Robert L Stevenson (1994b). The effectiveness of locator maps in increasing reader understanding of the geography of foreign news. *Journalism & Mass Communication Quarterly*, 71(4) : 937-46.

Jeffrey L Griffin and Robert L Stevenson (1996). The influence of statistical graphics on newspaper reader recall. *Visual Communication Quarterly*, 3(3) : 9-11.

Jerrold Griggs, Charles E Killian, and Carla D Savage (2004). Venn diagrams and symmetric chain decompositions in the Boolean lattice. *The Electronic Journal of Combinatorics*, 11(#R2) : 1-30.

Branko Grünbaum (1992). Venn Diagrams II. *Geombinatorics*, 2(2) : 25-32.

Stéphane Guindon and Olivier Gascuel (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5) : 696-704.

Corin Gurr and Konstantinos Tourlas (2000). Towards the principled design of software engineering diagrams. In: *Proceedings of the 22nd International Conference on Software Engineering (ICSE)*, pp. 509-18, Limerick, Ireland. IEEE.

Corin Gurr (2001). Aligning syntax and semantics in formalisations of visual languages. In: *Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments (HCC)*, pp. 60-61, Stresa, Italy. IEEE.

Ronny Hadany and David Harel (2001). A multi-scale algorithm for drawing graphs nicely. *Discrete Applied Mathematics*, 113(1) : 3-21.

M Haetzman, A M Elliott, B H Smith, P Hannaford, and W A Chambers (2003). Chronic pain and the use of conventional and alternative therapy. *Family Practice*, 20(2) : 147-54.

Sandra Hale, Joel Myerson, Lisa J Emery, Bonnie M Lawrence, and Carolyn DuFault (2007). Variation in working memory across the life span. In Andrew RA Conway, Christopher Jarrold, Michael J Kane, Akira Miyake, and John N Towse (eds.), *Variation in working memory,* Oxford University Press, New York, NY, USA, pp. 194-224.

Gary Hall (2011). *Digitize Me, Visualize Me, Search Me: Open Science and its Discontents*, Living Books About Life, Open Humanities Press.

David Harel and Yehuda Koren (2001). A fast multi-scale method for drawing large graphs. In: *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000), Lecture Notes in Computer Science 1984*, pp. 183-96, Colonial Williamsburg, VA, USA. Springer.

Mark Harrower and Cynthia Brewer (2003). Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1) : 27-37.

Robert Harvey, Eugene Schuster, and Barbara H Jennings (2013). Pleiohomeotic Interacts with the Core Transcription Elongation Factor Spt5 to Regulate Gene Expression in Drosophila. *PLOS ONE*, 8(7) : e70184.

Reid Hastie and Robyn M Dawes (2001). Thinking Rationally About Uncertainty. *Rational Choice in an Uncertain World: The Psychology of Judgement and Decision Making,* Sage Publications, Thousand Oaks, CA, USA.

Sarah T Hawley, Brian Zikmund-Fisher, Peter Ubel, Aleksandra Jancovic, Todd Lucas, and Angela Fagerlin (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, 73: 448-55.

Christopher G Healey, Kellogg S Booth, and James T Enns (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(2) : 107-35.

Christopher G Healey and James T Enns (2002). Perception and Painting: A search for effective, engaging visualizations. *Computer Graphics and Applications*, 22(2) : 10-15.

Thomas L Heath (1956). *The Thirteen Books of Euclid's Elements (3 vols.)*, Dover Publications, New York, NY, USA.

Jeffrey Heer and Michael Bostock (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In: *Proceedings of the 28th SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 203-12, Atlanta, GA, USA. ACM.

Jeremy Heil (2004). *vennX: Proportional Venn Diagrams*, Matlab Central. The MathWorks Inc. Accessed February 28, 2013, http://www.mathworks.com/matlabcentral/fileexchange/6116-proportional-venn-diagrams.

Julie Heiser, Doantam Phan, Maneesh Agrawala, Barbara Tversky, and Pat Hanrahan (2004). Identification and validation of cognitive design principles for automated generation of assembly instructions. In: *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pp. 311-19, Gallipoli, Italy. ACM.

David W Henderson (1963). Venn diagrams for more than four classes. *The American Mathematical Monthly*, 70(4) : 424-26.

Nathalie Henry and Jean-Daniel Fekete (2006). Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5) : 677-84.

Nathalie Henry, Jean-Daniel Fekete, and Michael J McGuffin (2007). NodeTrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6) : 1302-9.

Robin K Henson (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist*, 34(5) : 601-29.

Don Herbison-Evans (1995). Solving Quartics and Cubics for Graphics. In Alan W Paeth (ed.), *Graphics Gems V,* Morgan Kaufmann, San Francisco, CA, USA, pp. 3-15.

Robert Hess and David Field (1999). Integration of contours: new insights. *Trends in Cognitive Sciences*, 3(12) : 480-86.

Mary B Hesse (1966). *Models and analogies in science*, vol. 7, University of Notre Dame Press, Notre Dame, IN, USA.

Kenneth J Hill (1995). Matrix-based Ellipse Geometry. In Alan W Paeth (ed.), *Graphics Gems V,* Morgan Kaufmann, San Francisco, CA, USA, pp. 72-77.

Wendy Hilton (1981). *Dance of Court and Theater: The French Noble Style, 1690-1725*, Princeton, London, UK.

MHW Hobbs and Peter Rodgers (1998). Representing space: A hybrid genetic algorithm for aesthetic graph layout. In: *Proceedings of the 2nd International Workshop on Frontiers in Evolutionary Algorithms (FEA) at the 4th Joint Conference on Information Sciences (JCIS)*, pp. 415-18, Research Triangle Park, NC, USA.

Ulrich Hoffrage and Gerd Gigerenzer (1998). Using Natural Frequencies to Improve Diagnostic Inferences. *Academic Medicine*, 73(5) : 538-40.

Ulrich Hoffrage, Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer (2000). Communicating Statistical Information. *Science*, 290(5500) : 2261-62.

Ulrich Hoffrage, Gerd Gigerenzer, Stefan Krauss, and Laura Martignon (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, 84(3) : 343-52.

Nigel Holmes (2013). *Explanation Graphics*. Accessed March 30, 2013, http://nigelholmes.com.

Jana Holsanova, Henrik Rahm, and Kenneth Holmqvist (2006). Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. *Visual Communication*, 5(1) : 65-93.

Danny Holten and Jarke J van Wijk (2009). Force-Directed Edge Bundling for Graph Visualization. *Computer Graphics Forum*, 28(3) : 983-90.

Keith J Holyoak and Paul Thagard (1995). *Mental leaps: Analogy in creative thought*, MIT Press, Cambridge, MA, USA.

Wan-Jen Hong, Robert Tibshirani, and Gilbert Chu (2009). Local false discovery rate facilitates comparison of different microarray experiments. *Nucleic Acids Research*, 37(22) : 7483-97.

Adrian A Hopgood (2012). *Intelligent systems for engineers and scientists, 3rd Edition*, CRC Press, Boca Raton, FL, USA (hopgood2012intelligent).

John J Horton, David G Rand, and Richard J Zeckhauser (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3) : 399-425.

John Joseph Horton and Lydia B Chilton (2010). The labor economics of paid crowdsourcing. In: *Proceedings of the 11th ACM Conference on Electronic Commerce*, Cambridge, MA, USA.

Donald House, Alethea Bair, and Colin Ware (2005). On the Optimization of Visualizations of Complex Phenomena. In: *Proceedings of the 16th IEEE Visualization Conference (VIS)*, pp. 87-94, Minneapolis, MN, USA. IEEE Computer Society.

John Howse, Fernando Molina, John Taylor, Stuart Kent, and Joseph Yossi Gil (2001). Spider diagrams: A diagrammatic reasoning system. *Journal of Visual Languages & Computing*, 12(3) : 299-324.

John Howse, Gem Stapleton, and John Taylor (2005). Spider Diagrams. *London Mathematical Society (LMS) Journal of Computation and Mathematics*, 8: 145-94.

John Howse, Gem Stapleton, Kerry Taylor, and Peter Chapman (2011). Visualizing ontologies: A case study. In: *Proceedings of the 10th International Semantic Web Conference (ISWC)*, pp. 257-72, Bonn, Germany. Springer.

Yifan Hu (2005). Efficient and High-Quality Force-Directed Graph Drawing. *Mathematica Journal*, 10(1) : 37-71.

Yifan Hu (2011). Algorithms for Visualizing Large Networks. *Combinatorial Scientific Computing*.

Weidong Huang, Seok-Hee Hong, and Peter Eades (2006). Predicting graph reading performance: a cognitive approach. In: *Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation (APVis), vol. 60*, pp. 207-16, Tokyo, Japan. Australian Computer Society Inc.

Weidong Huang, Peter Eades, and Seok-Hee Hong (2008). Beyond time and error: a cognitive approach to the evaluation of graph drawings. In: *Proceedings of the 2008 conference on BEyond time and errors: novel evaLuation methods for Information Visualization, article no. 3*, Florence, Italy. ACM.

Weidong Huang, Peter Eades, and Seok-Hee Hong (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3) : 139-52.

Weidong Huang (2013). Establishing aesthetics based on human graph reading behavior: two eye tracking studies. *Personal and Ubiquitous Computing*, 17(1) : 93-105.

Julie A Huber, Hilary G Morrison, Susan M Huse, Phillip R Neal, Mitchell L Sogin, and Welch Mark, David B (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environmental Microbiology*, 11(5) : 1292-302.

Gary B Hughes and Mohcine Chraibi (2011). *Calculating Ellipse Overlap Area.* arXiv:1106.3787 [physics.comp-ph].

Hyun-Joo Lee Huh (1993). The Effect of Newspaper Picture Size on Readers' Attention, Recall, and Comprehension of Stories. In: *Proceedings of the 76th Annual Meeting of the Association for Education in Journalism and Mass Communication*, Kansas City, MO, USA. ERIC.

Tim Hulsen, Vlieg De, Jacob, and Wynand Alkema (2008). BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, 9: 488.

Leo M Hurvich (1981). *Color vision*, Sinauer Associates, Sunderland, MA, USA.

Shin'ichi Ichikawa (1989). The role of isomorphic schematic representation in the comprehension of counterintuitive Bayesian problems. *Journal of Mathematical Behavior*, 8: 269-81.

Eugene Isaacson (1994). *Analysis of numerical methods*, Courier Dover Publications, Mineola, NY, USA.

Ioan James (2002). *Remarkable Mathematicians: From Euler to von Neumann*, Cambridge University Press, Cambridge, UK.

Chris John (2005). Measuring and reducing clutter in Euler diagrams. *Proceedings of the 1st International Workshop on Euler Diagrams (Euler Diagrams 2004), Electronic Notes in Theoretical Computer Science vol. 134*, : 103-26.

Chris John, Andrew Fish, John Howse, and John Taylor (2006). Exploring the Notion of 'Clutter'in Euler Diagrams. In: *Proceedings of the 4th International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 4045*, pp. 267-82, Stanford, CA, USA. Springer.

Philip Nicholas Johnson-Laird, Paolo Legrenzi, Vittorio Girotto, Maria Sonino Legrenzi, and Jean-Paul Caverni (1999). Naive Probability: A Mental Model Theory of Extensional Reasoning. *Psychological Review*, 106(1) : 62-88.

Bela Julesz and James R Bergen (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *The Bell System Technical Journal*, 62(6).

Daniel Kahneman, Paul Slovic, and Amos Tversky (1982). *Judgment under uncertainty: Heuristics and biases*, Cambridge University Press, Cambridge, UK.

Tomihisa Kamada and Satoru Kawai (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1) : 7-15.

Yun Kang, Michael H Norris, Jan Zarzycki-Siek, William C Nierman, Stuart P Donachie, and Tung T Hoang (2011). Transcript amplification from single bacterium for transcriptome analysis. *Genome research*, 21(6) : 925-35.

Gaetano Kanizsa and Walter Gerbino (1976). Convexity and symmetry in figure-ground organization. In Mary Henle (ed.), *Vision and Artifact,* Springer Publishing Company, New York, NY, USA, pp. 25-32.

Kazutaka Katoh and Hiroyuki Toh (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, 26(15) : 1899-900.

Michael Kaufmann and Dorothea Wagner (2001). *Drawing graphs: methods and models*, vol. 2025, Springer Verlag, Berlin Heidelberg, Germany.

Vince Kellen, Susy Chan, and Xiaowen Fang (2007). Facilitating Conditional Probability Problems with Visuals. In: *Proceedings of the 12th International Conference on Human-Computer Interaction (HCI International), Interaction Platforms and Techniques (Part II), Lecture Notes in Computer Science 4551*, pp. 63-71, Beijing, China. Springer.

Vince Kellen, Susy Chan, and Xiaowen Fang (2013). Improving User Performance in Conditional Probability Problems with Computer-Generated Diagrams. In: *Proceedings of the 18th International Conference on Human-Computer Interaction (HCI International), Lecture Notes in Computer Science 8006*, pp. 183-92, Las Vegas, NV, USA.

Vincent J Kellen (2012), *The Effects of Diagrams and Relational Complexity on User Performance in Conditional Probability Problems in a Non-Learning Context*, PhD thesis, College of Computing and Digital Media of DePaul University, Chicago, IL, USA.

René Keller, Claudia M Eckert, and P John Clarkson (2006). Matrices or node-link diagrams: which visual representation is better for visualising connectivity models? *Information Visualization*, 5(1) : 62-76.

Stephen Kent, Mary Elizabeth Kaiser, Susana E Deustua, J Smith, Saul Adelman, Sahar Allam, Brian Baptista, Ralph C Bohlin, James L Clem, Alex Conley, Jerry Edelstein, Jay Elias, Ian Glass, Arne Henden, Steve Howell, Randy A Kimble, Jeffrey W Kruk, Michael Lampton, Eugene A Magnier, Stephan R McCandliss, Warren Moos, Nick Mostek, Stuart Mufson, Terry D Oswalt, Saul Perlmutter, Carlos Allende Prieto, Bernard J Rauscher, Adam Riess, Abhijit Saha, Mark Sullivan, Nicholas Suntzeff, Alan Tokunaga, Douglas Tucker, Robert Wing, Bruce Woodgate, and Edward L Wright (2009). *Photometric calibrations for 21st century science*. arXiv:0903.2799v1 [astro-ph.CO].

Stuart Kent (1997). Constraint diagrams: visualizing invariants in object-oriented models. In: *Proceedings of the 12th ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA)*, pp. 327-41, Atlanta, GA, USA. ACM.

Hans A Kestler, André Müller, Thomas M Gress, and Malte Buchholz (2005). Generalized Venn diagrams: a new method of visualizing complex genetic set relations. *Bioinformatics*, 21(8) : 1592-95.

Hans A Kestler, André Müller, Johann M Kraus, Malte Buchholz, Thomas M Gress, Hongfang Liu, David W Kane, Barry R Zeeberg, and John N Weinstein (2008). VennMaster: Area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinformatics*, 9: 67.

Bohyoung Kim, Bongshin Lee, and Jinwook Seo (2007). Visualizing set concordance with permutation matrices and fan diagrams. *Interacting with Computers*, 19(5) : 630-43.

Robert F Klees, Roman M Salasznyk, Scott Vandenberg, Kristin Bennett, and George E Plopper (2007). Laminin-5 activates extracellular matrix production and osteogenic gene focusing in human mesenchymal stem cells. *Matrix biology*, 26(2) : 106-14.

Allen Klinger and Nikos A Salingaros (2000). A pattern measure. *Environment and Planning B: Planning and Design*, 27: 537-47.

Stephen G Kobourov (2012). Spring Embedders and Force Directed Graph Drawing Algorithms. *Computing Research Repository (CoRR)*, abs/1201.3011.

Jonathan J Koehler (1997). One in Millions, Billions and Trillions: Lessons from People v. Collins (1968) for People v. Simpson (1995). *Journal of Legal Education*, 47: 214-23.

Kurt Koffka (1935). *Principles of Gestalt Psychology*, Harcourt Brace, New York, NY, USA.

Nicholas Kong, Jeffrey Heer, and Maneesh Agrawala (2010). Perceptual Guidelines for Creating Rectangular Treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 16(6) : 990-98.

Granino Arthur Korn and Theresa Marie Korn (1968). *Mathematical Handbook for Scientists and Engineers*, McGraw Hill, New York, NY, USA.

Robert Kosara and Jock Mackinlay (2013). Storytelling: The Next Step for Visualization. *Computer*, 46(5) : 44-50.

Maria Kozhevnikov, Mary Hegarty, and Richard E Mayer (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, 20(1) : 47-77.

Maria Kozhevnikov, Stephen Kosslyn, and Jennifer Shephard (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, 33(4) : 710-26.

Gunther R Kress and Theo van Leeuwen (2006). *Reading Images: The Grammar of Visual Design, 2nd Edition*, Routledge, Padstow, UK.

James K Kroger, Fred W Sabb, Christina L Fales, Susan Y Bookheimer, Mark S Cohen, and Keith J Holyoak (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cerebral Cortex*, 12(5) : 477-85.

Jon A Krosnick, Fan Li, and Darrin R Lehman (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of personality and social psychology*, 59(6) : 1140-52.

Tevye R Krynski and Joshua B Tenenbaum (2007). The Role of Causality in Judgment Under Uncertainty. *Journal of Experimental Psychology: General*, 136(3) : 430-50.

Martin Krzywinski, Jacqueline Schein, İnanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco A Marra (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, 19(9) : 1639-1645. http://www.circos.ca.

Elke Kurz-Milcke, Gerd Gigerenzer, and Laura Martignon (2008). Transparency in Risk Communication: Graphical and Analog Tools. *Annals of the New York Academy of Sciences*, 1128(1) : 18-28.

Sr de La Cuisse (1762). *Le Répertoire des bals ou théorie-pratique des contredanses, décrites d'une maniere aisée avec des figures démonstratives pour les pouvoir danser facilement, auxquelles on a ajouté les airs notés*, Cailleau & Mlle Castagnery, Paris, France.

Gerald J Laabs and Herbert George Baker (1989). Selection of critical tasks for Navy job performance measures. *Military Psychology*, 1(1) : 3-16.

Oliver Lenz and Alessia Fornoni (2006). Chronic kidney disease care delivered by US family medicine and internal medicine trainees: results from an online survey. *BMC Medicine*, 4: 30.

Li Liao, Xue-Wei Xu, Xia-Wei Jiang, Chun-Sheng Wang, Dong-Sheng Zhang, Jian-Yu Ni, and Min Wu (2011). Microbial diversity in deep-sea sediment from the cobalt-rich crust deposit region in the Pacific Ocean. *FEMS Microbiology Ecology*, 78(3) : 565-85.

Lee P Lim, Nelson C Lau, Philip Garrett-Engele, Andrew Grimson, Janell M Schelter, John Castle, David P Bartel, Peter S Linsley, and Jason M Johnson (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027) : 769-73.

Dennis V Lindley (1994). Making Decisions, 2nd Edition.

Isaac M Lipkus, Greg Samsa, and Barbara K Rimer (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21: 37-44.

Isaac M Lipkus (2007). Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Medical Decision Making*, 27(5) : 696-713.

M L Lister, M F Aller, H D Aller, D C Homan, K I Kellermann, Y Y Kovalev, A B Pushkarev, J L Richards, E Ros, and T Savolainen (2013). MOJAVE. X. Parsec-scale Jet Orientation Variations and Superluminal Motion in Active Galactic Nuclei. *The Astronomical Journal*, 146(5) : 120.

Kyle Littlefield and Mattew Monroe (2013). *Venn Diagram Plotter*. PNNL and OMICS.PNL.GOV, Richland, WA, USA. Accessed January 28, 2014, http://omics.pnl.gov/software/VennDiagramPlotter.php.

Zhicheng Liu and John Stasko (2010). Theories in information visualization: What, why and how. In: *Proceedings of the IEEE VisWeek Workshop on the Role of Theory in Information Visualization*, Salt Lake City, UT, USA.

Zili Liu, David W Jacobs, and Ronen Basri (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Research*, 39(25) : 4244-57.

Ramon Lull (1617). *Opera ea quae ad adinventam ab ipso artem universalem scientiarum artiumque omnium brevi compendio, firmaque memoria apprehendendarum, locupletissimaque vel oratione ex tempore pertractandarum, pertinent*, Sumptibus Lazari Zetzneri.

Laura Macchi (2000). Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes*, 82(2) : 217-36.

Michael Macdonald-Ross (1977). How Numbers are Shown. *Educational Technology Research and Development*, 25(4) : 359-409.

Alan M. MacEachren (2004). *How Maps Work: Representation, Visualization, and Design*, The Guilford Press, New York, NY, USA.

Jock Mackinlay (1986a), *Automatic Design of Graphical Presentations*, PhD thesis, Stanford University, Stanford, CA, USA.

Jock Mackinlay (1988). Applying a theory of graphical presentation to the graphic design of user interfaces. In: *Proceedings of the 1st annual ACM SIGGRAPH Symposium on User Interface Software (UIST)*, pp. 179-89, New York, NY, USA. ACM.

Jock Mackinlay (2009). Designing Great Visualizations. In: *Web document, accessed: November 19, 2012*, Tableau Software. http://www.tableausoftware.com/whitepapers/designing-great-visualizations.

Jock Mackinlay (1986b). Automating the Design of Graphical Presentations of Relational Information. *ACM Transaction on Graphics*, 5(2) : 110-41.

Matthieu Maitre and Ha T Nguyen (2004). *Making Videos from Images of Non-Rigid Objects*. Accessed July 15, 2013, http://www.ifp.illinois.edu/~hanguyen/research/final_report.pdf.

Khalegh Mamakani, Wendy Myrvold, and Frank Ruskey (2011). Generating all simple convexly-drawable polar symmetric 6-Venn diagrams. *Combinatorial Algorithms, Lecture Notes in Computer Science*, 7056: 275-86.

Karlheinz Mann and Matthias Mann (2011). In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. *Proteome Science*, 9: 7.

Dinesh Manocha and James Demmel (1994). Algorithms for intersecting parametric and algebraic curves I: simple intersections. *ACM Transactions on Graphics*, 13(1) : 73-100.

R Timothy Marler and Jasbir S Arora (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6) : 369-95.

Roger J Marshall (2005). Scaled rectangle diagrams can be used to visualize clinical and epidemiological data. *Journal of Clinical Epidemiology*, 58(10) : 974-81.

Bronwen Martin, Wayne Chadwick, Tie Yi, Sung-Soo Park, Daoyuan Lu, Bin Ni, Shekhar Gadkaree, Kathleen Farhang, Kevin G Becker, and Stuart Maudsley (2012). VENNTURE—A Novel Venn Diagram Investigational Tool for Multiple Pharmacological Dataset Analysis. *PLoS ONE*, 7(5) : e36911.

Michael A Martin (2003). It's like… you know: The Use of Analogies and Heuristics in Teaching Introductory Statistical Methods. *Journal of Statistics Education*, 11(2).

Sergio Cesare Masin (1997). The luminance conditions of transparency. *Perception*, 26: 39-50.

Hans-Joachim Meihoefer (1969). The utility of the circle as an effective cartographic symbol. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 6(2) : 105-17.

Hans-Joachim Meihoefer (1973). The visual perception of the circle in thematic maps/experimental results. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(1) : 63-84.

Barbara A Mellers and A Peter McGraw (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995). *Psychological Review*, 106(2) : 417-24.

Fabio Metelli (1974). The perception of transparency. *Scientific American*, 230(4) : 90-98.

Luana Micallef and Peter Rodgers (2009). Poster: Force-Directed Layout for Euler Diagrams. In: *Compendium of the IEEE Information Visualization (InfoVis)*, Atlantic City, NJ, USA. IEEE Computer Society. http://www.eulerdiagrams.org/eulerForce.

Luana Micallef (2012). Visualizations with Venn and Euler Diagrams. In: *Proceedings of the 12th Annual Grace Hopper Celebration of Women in Computing (GHC), PhD Forum*, Baltimore, MA, USA.

Luana Micallef, Pierre Dragicevic, and Jean-Daniel Fekete (2012). Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics*, 18(12) : 2536-45.

Luana Micallef and Peter Rodgers (2012). Poster: Drawing Area-Proportional Venn-3 Diagrams Using Ellipses. In: *12th Annual Grace Hopper Celebration of Women in Computing (GHC), Poster Session and ACM Student Research Competition (SRC)*, Baltimore, MD, USA. http://www.eulerdiagrams.org/eulerAPE.

Luana Micallef and Peter Rodgers (2014a). eulerAPE: Drawing Area-Proportional 3-Venn Diagrams Using Ellipses. *PLoS ONE*, 9(7) : e101717.

Luana Micallef and Peter Rodgers (2014b). eulerForce: Force-directed Layout for Euler Diagrams. *Journal of Visual Languages and Computing*, 10.1016/j.jvlc.2014.09.002, in press.

Luana Micallef and Peter Rodgers (2014c). Computing the Region Areas of Euler Diagrams Drawn with Three Ellipses. In: *Proceedings of the 4th International Workshop on Euler Diagrams, CEUR-WS. org vol. 1244*, pp. 1-15, Melbourne, Australia.

George A. Miller (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63(2) : 81-97.

Robert B Miller (1968). Response time in man-computer conversational transactions. In: *Proceedings of the December 9-11, 1968 (AFIPS) fall joint computer conference, part I*, pp. 267-77, San Francisco, CA, USA. ACM.

Charles Joseph Minard (1869). Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813. *Des Tableaux Graphiques et Cartes Figuratives de M. Minard (1845-1869), Item 28,* Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris, France.

Melanie Mitchell, John H Holland, and Stephanie Forrest (1993). When will a genetic algorithm outperform hill climbing? In: *Proceedings of the 7th International Conference on Neural Information Processing Systems (NIPS), vol. 6*, pp. 51-58, Denver, CO, USA.

Daniel R Montello (2002). Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. *Cartography and Geographic Information Science*, 29(3) : 283-304.

Angela Morelli (2012a). *Virtual Water - Discover how much WATER we EAT everyday*, The Water We Eat: an infographic story. Accessed March 29, 2013, http://www.angelamorelli.com/water.

Angela Morelli (2012b), 'The Water We Eat', *Che Futuro*, September 24, 2012b, sec. Social Innovation (also at http://www.angelamorelli.com/1/?projects=virtual-water-an-infographic-story).

Angela Morelli (2013). *The Water We Eat: an infographic story*. Accessed March 29, 2013, http://www.angelamorelli.com/1/?projects=virtual-water-an-infographic-story.

Philippe Moret, Walter Binder, Alex Villazón, Danilo Ansaloni, and Abbas Heydarnoori (2010). Visualizing and exploring profiles with calling context ring charts. *Software: Practice and Experience*, 40(9) : 825-47.

John Morgan (2012), 'Between systems, one dead, one stuck in legislative limbo', *Times Higher Education*, May 17, 2012, (available online at http://www.timeshighereducation.co.uk/419955.article).

Rodrigo Moro (2010). Sets or frequencies?: how to explain the facilitation effect in conditional probability problems. *Anales de Psicología*, 26(1) : 181-88.

Frederick Mosteller (1987). *Fifty challenging problems in probability with solutions*, Dover Publications, Mineola, NY, USA.

Robert S Moyer and Thomas K Landauer (1967). Time required for judgements of numerical inequality. 215: 1519-20.

Kathy T Mullen (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1) : 381-400.

Alena Myšičková and Martin Vingron (2012). Detection of interacting transcription factors in human tissues using predicted DNA binding affinity. *BMC Genomics*, 13(Suppl 1) : S2.

Robin Naidoo and Takuya Iwamura (2007). Global-scale mapping of economic benefits from agricultural lands: implications for conservation priorities. *Biological Conservation*, 140(1) : 40-49.

National Geographic (2009), 'The Carbon Bathtub', *National Geographic*, December, 2009, (graphic: Nigel Holmes; sources: John Sterman and David Archer; graphic available at http://ngm.nationalgeographic.com/big-idea/05/carbon-bath).

David E Nelson, Bradford W Hesse, and Robert T Croyle (2009). Making Data Talk: Communicating Public Health Data to the Public, Policy Makers and the Press.

Wendy Nelson, Valerie F Reyna, Angela Fagerlin, Isaac Lipkus, and Ellen Peters (2008). Clinical Implications of Numeracy: Theory and Practice. *Annals of Behavioral Medicine*, 35(3) : 261-74.

Marie Neurath and Robin Kinross (2009). *The transformer: principles of making Isotype charts*, Hyphen Press, London, UK.

Otto Neurath (1936). *International Picture Language*, Kegan Paul, Trench, Trubner & Co, London, UK.

Otto Neurath (2010). *From hieroglyphics to Isotype: a visual autobiography*, Hyphen Press, London, UK.

Otto Neurath and Marie Neurath (2012). *Isotype Institute*. Accessed March 23, 2013, www.fulltable.com/iso.

Isaac Newton (1687). *Principia*, Jussu Societatis Regiæ ac Typis Josephi Streater, Prostat apud plures bibliopolas, London, UK.

Isaac Newton (1704). *Opticks: or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light*, Printed for Sam Smith and Benj Walford, printers to the Royal Society, at the Prince's Arms in St Paul's Church-yard, London, UK.

Austin Lee Nichols and Jon K Maner (2008). The good-subject effect: investigating participant demand characteristics. *The Journal of general psychology*, 135(2) : 151-65.

Florence Nightingale (1858). *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army: Founded Chiefly on the Experience of the Late War. By Florence Nightingale. Presented by Request to the Secretary of State for War*, Harrison and Sons, London, UK.

F Nikulenkov, C Spinnler, H Li, C Tonelli, Y Shi, M Turunen, T Kivioja, I Ignatiev, A Kel, J Taipale, and G Selivanova (2012). Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. *Cell Death & Differentiation*, 19: 1992-2002.

Takao Nishizeki and Md Saidur Rahman (2004). *Planar Graph Drawing*, Lecture Notes Series on Computing, vol. 12, World Scientific Publishing Co, Singapore.

Terence RF Nonweiler (1968). Algorithm 326: Roots of low-order polynomial equations. *Communications of the ACM*, 11(4) : 269-70.

Donald A Norman (1988). *The Psychology of Everyday Things*, Basic Books, New York, NY, USA.

Donald A Norman (1991). Cognitive Artifacts. In John M. Carroll (ed.), *Designing Interaction,* Cambridge University Press, New York, NY, USA, pp. 17-38.

Donald A Norman (2002). *The Design of Everyday Things*, Basic Books, New York, NY, USA.

Donald A Norman (2004). *Emotional design: Why we love (or hate) everyday things*, Basic Books, New York, NY, USA.

Laura R Novick and Keith J Holyoak (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(3) : 398-415.

Natalie A Obrecht, Britta Anderson, Jay Schulkin, and Gretchen B Chapman (2012). Retrospective Frequency Formats Promote Consistent Experience-Based Bayesian Judgments. *Applied Cognitive Psychology*, 26(3) : 436-40.

Natalie A Obrecht and Dana L Chesney (2013). Sample representativeness affects whether judgments are influenced by base rate or sample size. *Acta Psychologica*, 142(3) : 370-82.

T Ohrlander, M Dencker, and S Acosta (2011). Preoperative Spirometry Results as a Determinant for Long-term Mortality after EVAR for AAA. *European Journal of Vascular and Endovascular Surgery*, 43(1) : 43-47.

Juan Carlos Oliveros (2007). *VENNY: An interactive tool for comparing lists with Venn Diagrams*. BioinfoGP, CNB-CSIC. Accessed February 28, 2013, http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Daniel M Oppenheimer, Tom Meyvis, and Nicolas Davidenko (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4) : 867-72.

John Page (2009). *Ellipses*. Math Open Reference. Accessed July 10, 2013, http://www.mathopenref.com/tocs/ellipsetoc.html.

Alexandre Palagi, Jennifer Koh, Mathieu Leblanc, David Wilson, Sébastien Dutertre, Glenn F King, Graham M Nicholson, and Pierre Escoubas (2013). Unravelling the complex venom landscapes of lethal Australian funnel-web spiders (Hexathelidae: Atracinae) using LC-MALDI-TOF mass spectrometry. *Journal of Proteomics*, in press.

Stephen E Palmer (1992). Common region: A new principle of perceptual grouping. *Cognitive Psychology*, 24(3) : 436-47.

Alex T Pang, Craig M Wittenbrink, and Suresh K Lodha (1997). Approaches to uncertainty visualization. *The Visual Computer*, 13(8) : 370-90.

Gabriele Paolacci, Jesse Chandler, and Panagiotis Ipeirotis (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5) : 411-19.

Steve Pasternack and Sandra H Utt (1990). Reader Use and Understanding of Newspaper Informational Graphics. *Newspaper Research Journal*, 11(2) : 28-41.

Adam Perer (2010). Finding Beautiful Insights in the Chaos of Social Network Visualizations. In Julie Steele and Noah Iliinsky (eds.), *Beautiful Visualization: Looking at Data through the Eyes of Experts,* O'Reilly Media, Sebastopol, CA, USA, pp. 157-74.

Howard Perlman (2013). *How much water is on Earth?*, Water Science Photo Gallery. U.S. Department of the Interior, U.S. Geological Survey. Accessed May 5, 2013, http://ga.water.usgs.gov/edu/2010/gallery/global-water-volume.html.

Federico Pernici (2005), *Two Results in Computer Vision using Projective Geometry*, PhD thesis, University of Florence, Florence, Italy.

Ellen Peters, Daniel Västfjäll, Paul Slovic, C K Mertz, Ketti Mazzocco, and Stephan Dickert (2006). Numeracy and decision making. *Psychological Science*, 17(5) : 407-13.

Ellen Peters, Judith Hibbard, Paul Slovic, and Nathan Dieckmann (2007). Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs*, 26(3) : 741-48.

Ellen Peters (2012). Beyond Comprehension The Role of Numeracy in Judgments and Decisions. *Current Directions in Psychological Science*, 21(1) : 31-35.

Steven Pinker (1994). *The Language Instinct: The New Science of Language and Mind*, The Penguin Press, London, UK.

Peter Pirolli and Stuart Card (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: *Proceedings of the 1st International Conference on Intelligence Analysis (IA)*, McLean, VA, USA.

Mehdi Pirooznia, Vijayaraj Nagarajan, and Youping Deng (2007). GeneVenn - A web application for comparing gene lists using Venn diagrams. *Bioinformation*, 1(10) : 420-22.

William Playfair (1786). Commercial and political atlas: Representing, by copper-plate charts, the progress of the commerce, revenues, expenditure, and debts of England, during the whole of the eighteenth century.

William Playfair (1801). *The Statistical Breviary: Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe; Illustrated with Stained Copper-plate Charts the Physical Powers of Each Distinct Nation with Ease and Perspicuity: to which is Added, a Similar Exhibition of the Ruling Powers of Hindoostan*, T Bensley, London, UK.

Alexei A Podtelezhnikov (2008). *Venn diagram of amino acid properties*, Proteins and DNA. Accessed April 1, 2013, https://sites.google.com/site/apodtele/aa_venn_diagram.jpg.

E C Poulton and P R Freeman (1966). Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 66(1) : 1-8.

Melanie Price, Rachel Cameron, and Phyllis Butow (2007). Communicating risk information: the influence of graphical display format on quantitative information perception—accuracy, comprehension and preferences. *Patient Education and Counseling*, 69(1-3) : 121-28.

Brian A Primack, Ariel Shensa, Kevin H Kim, Mary V Carroll, Mary T Hoban, E Victor Leino, Thomas Eissenberg, Kathleen H Dachille, and Michael J Fine (2013). Waterpipe smoking among US university students. *Nicotine & Tobacco Research*, 15(1) : 29-35.

Helen C Purchase, Jo-Anne Allder, and David Carrington (2001). User preference of graph layout aesthetics: A UML study. In: *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000), Lecture Notes in Computer Science 1984*, pp. 5-18, Colonial Williamsburg, VA, USA. Springer.

Helen C Purchase (2002). Metrics for Graph Drawing Aesthetics. *Journal of Visual Languages & Computing*, 13(5) : 501-16.

Helen C Purchase, David Carrington, and Jo-Anne Allder (2002). Empirical evaluation of aesthetics-based graph layout. *Empirical Software Engineering*, 7(3) : 233-55.

Helen C Purchase, Eve Hoggan, and Carsten Görg (2007). How important is the "mental map"?--an empirical investigation of a dynamic graph layout algorithm. In: *Proceedings of the 14th International Symposium on Graph Drawing (GD 2006), Lecture Notes in Computer Science 4372*, pp. 184-95, Karlsruhe, Germany. Springer.

Helen C Purchase, Christopher Pilcher, and Beryl Plimmer (2012). Graph drawing aesthetics—created by users, not algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 18(1) : 81-92.

Helen C. Purchase (2000). Effective information visualisation: a study of graph drawing aesthetics and algorithms. *Interacting with Computers*, 13(2) : 147-62.

Tom Quick, Chrystopher Nehaniv, Kerstin Dautenhahn, and Graham Roberts (2006). Sensorimotor information flow in genetic regulatory network driven control systems. In: *Proceedings of the 10th International Conference on the Simulation and Synthesis of Living Systems (Artificial Life)*, pp. 351-57, Bloomington, IN, USA.

Jill L Quilici and Richard E Mayer (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, 88(1) : 144-61.

Jyotika Ramaprasad (1991). Informational Graphics in Newspapers. *Newspaper Research Journal*, 12: 92-101.

Christian RA Regenbrecht, Marc Jung, Hans Lehrach, and James Adjaye (2008). The molecular basis of genistein-induced mitotic arrest and exit of self-renewal in embryonal carcinoma and primary cancer cell lines. *BMC Medical Genomics*, 1: 49.

Robert JD Reid, Sergio González-Barrera, Ivana Sunjevaric, David Alvaro, Samantha Ciccone, Marisa Wagner, and Rodney Rothstein (2011). Selective ploidy ablation, a high-throughput plasmid transfer protocol, identifies new genes affecting topoisomerase I–induced DNA damage. *Genome Research*, 21(3) : 477-86.

Alexander Renkl, Robin Stark, Hans Gruber, and Heinz Mandl (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary educational psychology*, 23(1) : 90-108.

Valerie F Reyna, Wendy L Nelson, Paul K Han, and Nathan F Dieckmann (2009). How Numeracy Influences Risk Comprehension and Medical Decision Making. *Psychological bulletin*, 135(6) : 943-73.

Nathalie Henry Riche and Tim Dwyer (2010). Untangling Euler Diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 16(6) : 1090-99.

Andrew Robinson (2007). *The Story of Writing: Alphabets, Hieroglyphs & Pictographs*, Thames & Hudson, London, UK.

Peter Rodgers, Leishi Zhang, and Andrew Fish (2008a). General Euler Diagram Generation. In: *Proceedings of the 5th International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 5223*, pp. 13-27, Herrsching, Germany. Springer.

Peter Rodgers, Leishi Zhang, Gem Stapleton, and Andrew Fish (2008b). Embedding Wellformed Euler diagrams. In: *Proceedings of the 12th International Conference on Information Visualization (IV)*, pp. 585-93, London, UK. IEEE Computer Society.

Peter Rodgers, Jean Flower, Gem Stapleton, and John Howse (2009). Some results for drawing area proportional Venn3 with convex curves. In: *Proceedings of the 13th International Conference on Information Visualization (IV)*, pp. 667-72, Barcelona, Spain. IEEE Computer Society.

Peter Rodgers, Jean Flower, Gem Stapleton, and John Howse (2010a). Drawing Area-Proportional Venn-3 Diagrams with Convex Polygons. In: *Proceedings of the 6th International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 6170*, pp. 54-68, Portland, OR, USA. Springer.

Peter Rodgers, Gem Stapleton, John Howse, and Leishi Zhang (2010b). Euler Graph Transformations for Euler Diagram Layout. In: *Proceedings of the 27th IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 111-18, Leganés-Madrid, Spain. IEEE.

Peter Rodgers, Jean Flower, and Gem Stapleton (2012a). Introducing 3D Venn and Euler Diagrams. In: *Proceedings of the 3rd International Workshop on Euler Diagrams, CEUR-WS. org vol. 854*, pp. 92-106, Canterbury, UK.

Peter Rodgers, Leishi Zhang, and Helen Purchase (2012b). Wellformedness Properties in Euler Diagrams: Which Should Be Used? *IEEE Transactions on Visualization and Computer Graphics*, 18(7) : 1089-100.

Peter Rodgers, John Howse, Gem Stapleton, and Jean Flower (2014). Drawing Area-Proportional Euler Diagrams Representing Up To Three Sets. *IEEE Transactions on Visualization and Computer Graphics*, 20(1).

Fabiana Aparecida Rodrigues, Juliana Marcolino-Gomes, Josirlei de Fátima Corrêa Carvalho, Leandro Costa do Nascimento, Norman Neumaier, José Renato Bouças Farias, Marcelo Falsarella Carazzolle, Francismar Corrêa Marcelino, and Alexandre Lima Nepomuceno (2012). Subtractive libraries for prospecting differentially expressed genes in the soybean under water deficit. *Genetics and Molecular Biology*, 35(1) : 304-14.

Ruth Rosenholtz, Yuanzhen Li, Jonathan Mansfield, and Zhenlan Jin (2005). Feature Congestion: A Measure of Display Clutter. In: *Proceedings of the 23rd SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 761-70, Portland, OR, USA. ACM.

Richard E Rosenthal (1985). Concepts, Theory, and Techniques Principles of Multiobjective Optimization. *Decision Sciences*, 16(2) : 133-52.

Hans Rosling (2010). *Hans Rosling's 200 countries, 200 years, 4 minutes*, The Joy of Stats. British Broadcasting Corporation (BBC). Accessed March 30, 2013, http://www.bbc.co.uk/programmes/p00cgkfk.

Joel Ross, Lilly Irani, M Silberman, Andrew Zaldivar, and Bill Tomlinson (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In: *Proceedings of the 28th SIGCHI Conference on Human Factors in Computing Systems (CHI), Extended Abstracts*, pp. 2863-72, Atlanta, GA, USA. ACM.

Frank Ruskey and Mark Weston (1997). A Survey of Venn Diagrams. *Electronic Journal of Combinatoric*, 4: Dynamic Survey DS5 (revised in 2001 and 2005).

Frank Ruskey, Carla D Savage, and Stan Wagon (2006). The Search for Simple Symmetric Venn Diagrams. *Notices of the AMS*, 53(11) : 1304-11.

Timothy A Salthouse (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, 103(3) : 403-28.

Yuri Sato, Koji Mineshima, and Ryo Takemura (2010). The efficacy of Euler and Venn diagrams in deductive reasoning: Empirical findings. In: *Proceedings of the 6th International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 6170*, pp. 6-22, Portland, OR, USA. Springer.

Roger C Schank (1990). *Tell me a story: Narrative and intelligence*, Northwestern University Press, Evanston, IL, USA.

Marilyn M Schapira, Ann B Nattinger, and Colleen A McHorney (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*, 21(6) : 459-67.

Christoph Scheiner (1626). *Rosa Ursina sive Sol ex admirando facularum & macularum*, Bracciano: Andreas Phaeus (1626-1630).

Maximilian Schich (2010). Revealing Matrices. In Julie Steele and Noah Iliinsky (eds.), *Beautiful Visualization: Looking at Data through the Eyes of Experts,* O'Reilly Media, Sebastopol, CA, USA (http://revealingmatrices.schich.info).

James A Schirillo and Eric R Stone (2005). The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. *Risk Analysis*, 25(3) : 555-66.

Philip Schneider and David H Eberly (2002). *Geometric tools for computer graphics*, Morgan Kaufmann, San Francisco, CA, USA.

Emery Schubert (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology*, 51(3) : 154-65.

Hans-Jörg Schulz (2011). Treevis.net: A Tree Visualization Reference. *IEEE Computer Graphics and Applications*, 31(6) : 11-15.

David Schum (1991). Jonathan Cohen and Thomas Bayes on the analysis of chains of reasoning. *Probability and Rationality: Studies on L. Jonathan Cohen's Philosophy of Science*, 21: 99.

David A Schum (1999). Probabilistic reasoning and the science of complexity. In James Shanteau, Barbara A Mellers, and David A Schum (eds.), *Decision Science and Technology: Reflections on the Contributions of Ward Edwards,* Kluwer Academic Publishers, Norwell, MA, USA, pp. 183-210.

Miriam W Schustack and John R Anderson (1979). Effects of analogy to prior knowledge on memory for new information. *Journal of Verbal Learning and Verbal Behavior*, 18(5) : 565-83.

John Scott (2012). *Social Network Analysis, 3rd Edition*, SAGE Publications Ltd, London, UK (scott2012social).

Thomas W Sederberg and Scott R Parry (1986). Comparison of three curve intersection algorithms. *Computer-Aided Design*, 18(1) : 58-63.

Thomas W Sederberg and Tomoyuki Nishita (1990). Curve intersection using Bézier clipping. *Computer-Aided Design*, 22(9) : 538-49.

Peter Sedlmeier and Gerd Gigerenzer (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, 10(1) : 33-51.

Peter Sedlmeier (1999). *Improving Satistical Reasoning: Theoretical Models and Practical Implications*, Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Peter Sedlmeier and Gerd Gigerenzer (2001). Teaching Bayesian Reasoning in Less Than Two Hours. *Journal of Experimental Psychology: General*, 130(3) : 380-400.

Edward Segel and Jeffrey Heer (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6) : 1139-48.

Chris Seidel (2005). *Pangloss Venn Diagram Generator*. Accessed February 28, 2013, http://www.pangloss.com/seidel/Protocols/venn.cgi.

Steve Selvin (1975). On the Monty Hall problem. *American Statistician*, 29(3) : 134.

Matthias Shapiro (2004). Once Upon a Stacked Time Series: The Importance of Storytelling in Information Visualization. *Beautiful Visualization: Looking at Data through the Eyes of Experts,* O'Reilly Media, Sebastopol, CA, USA, pp. 15-36.

Aaron D Shaw, John J Horton, and Daniel L Chen (2011). Designing incentives for inexpert human raters. In: *Proceedings of the 14th ACM International Conference on Computer Supported Cooperative Work (CSCW)*, pp. 275-84, Hangzhou, China. ACM.

Yang Shen, Jicheng Hu, Yanan Lu, and Xiaofeng Wang (2012). Stock trends prediction by hypergraph modeling. In: *Proceedings of the 3rd IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 104-7, Beijing, China. IEEE.

SL Sheridan, M Pignone, and others (2002). Numeracy and the medical student's ability to interpret data. *Effective clinical practice: ECP*, 5(1) : 35-40.

Atsushi Shimojima (1996), *On the efficacy of representation*, PhD thesis, Indiana University, Bloomington, IN, USA.

Atsushi Shimojima (2004). Inferential and expressive capacities of graphical representations: Survey and some generalizations. In: *Proceedings of the 3rd International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 2980*, pp. 18-21, Cambridge, UK. Springer.

Shinsuke Shimojo and Shin'ichi Ichikawa (1989). Intuitive reasoning about probability: Theoretical and experimental analyses of the "problem of three prisoners". *Cognition*, 32(1) : 1-24.

Li (Stan) Shiqun (2009). Customized Proportional Venn Diagrams from SAS® Syste. In: *Proceedings of the 22nd NorthEast SAS Users Group Annual Conference (NESUG)*, Burlington, VT, USA.

Ben Shneiderman (1992). Tree visualization with Tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11(1) : 92-99.

Kirsty R Short, Marloes Vissers, Kleijn de, Stan, Albert L Zomer, Katherine Kedzierska, Emma Grant, Patrick C Reading, Peter W M Hermans, Gerben Ferwerda, and Dimitri A Diavatopoulos (2013). Bacterial Lipopolysaccharide Inhibits Influenza Virus Infection of Human Macrophages and the Consequent Induction of CD8+ T Cell Immunity. *Journal of Innate Immunity*, in press.

Paolo Simonetto, David Auber, and Daniel Archambault (2009). Fully automatic visualisation of overlapping sets. *Computer Graphics Forum*, 28(3) : 967-74.

Paolo Simonetto (2011), *Visualisation of Overlapping Sets and Clusters with Euler Diagrams*, PhD thesis, Ećole doctorale de mathématiques et informatique, Université Bordeaux 1, Talence, France.

Steven A Sloman, David Over, Lila Slovak, and Jeffrey M Stibel (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, 91(2) : 296-309.

Paul Slovic, Ellen Peters, Melissa L Finucane, and Donald G MacGregor (2005). Affect, risk, and decision making. *Health Psychology*, 24(4S) : S35-40.

Gordon Smyth (2005). Limma: linear models for microarray data. In Robert Gentleman, Vincent J Carey, Wolfgang Huber, Rafael A Irizarry, and Sandrine Dudoit (eds.), Bioinformatics and Computational Biology Solutions using R and Bioconductor, Springer, New York, NY, USA, pp. 397-420.

Deborah Sole and Daniel Gray Wilson (2002). *Storytelling in organizations: The power and traps of using stories to share knowledge in organizations*. Learning Innovations Laboratories (LILA), Harvard, Graduate School of Education, Cambridge, MA, USA. Accessed March 29, 2013, http://www.providersedge.com/docs/km_articles/Storytelling_in_Organizations.pdf.

Joan B Soriano, Kourtney J Davis, Bobbie Coleman, George Visick, David Mannino, and Neil B Pride (2003). The Proportional Venn Diagram of Obstructive Lung Disease*: Two Approximations From the United States and the United Kingdom. *CHEST*, 124(2) : 474-81.

Bettina Speckmann (2006). Algorithms for cartograms and other specialized maps. *Geo-Information and Computational Geometry*, 44: 26-34.

Bettina Speckmann, Marc Kreveld, and Sander Florisson (2006). A linear programming approach to rectangular cartograms. *Progress in Spatial Data Handling*, : 529-46.

Ian Spence and Forrest W Young (1978). Monte Carlo studies in nonmetric scaling. *Psychometrika*, 43(1) : 115-17.

Ian Spence (2006). William Playfair and the Psychology of Graphs. In: *Proceedings of the American Statistical Association, Section of Statistical Graphs*, pp. 2426-36, Alexandria, VA, USA. American Statistical Association.

Robert Spence (1996). Visualisation really has nothing to do with computers. In: *Proceedings of the 14th Annual Eurographics UK Conference (EGUK)*, pp. 1-8, London, UK. IEEE.

Robert Spence (2001). *Information Visualization*, Addison-Wesley, Harlow, England.

Robert Spence (2007). *Information Visualization: Design for Interaction, 2nd Edition*, Pearson Education, Essex, England.

David Spiegelhalter, Mike Pearson, and Ian Short (2011). Visualizing Uncertainty About the Future. *Science*, 333(6048) : 1393-400.

Gem Stapleton (2005). A survey of reasoning systems based on Euler diagrams. *Electronic Notes in Theoretical Computer Science*, 134: 127-51.

Gem Stapleton, Peter Rodgers, John Howse, and John Taylor (2007). Properties of Euler diagrams. In: *Proceedings of the Workshop on the Layout of (Software) Engineering Diagrams (LED), Electronic Communications of the EASST vol. 7*, Coeur d'Alène, ID, USA.

Gem Stapleton, Andrew Fish, and Peter Rodgers (2008a). Abstract Euler Diagram Isomorphism. In: *Proceedings of the 14th International Conference on Distributed Multimedia Systems, Visual Languages and Computing (DMS)*, Boston, MA, USA. Knowledge Systems Institute.

Gem Stapleton, John Howse, Peter Rodgers, and Leishi Zhang (2008b). Generating Euler Diagrams from Existing Layouts. In: *Proceedings of the 2nd Layout of (Software) Engineering Diagrams (LED), Electronic Communications of the EASST vol. 13*, Herrsching am Ammersee, Germany.

Gem Stapleton, John Howse, and Peter Rodgers (2010). A graph theoretic approach to general Euler diagram drawing. *Theoretical Computer Science*, 411(1) : 91-112.

Gem Stapleton and Peter Rodgers (2011). Drawing Euler Diagrams with Circles and Ellipses. In: *Proceedings of the 28th IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 209-12, Pittsburgh, PA, USA. IEEE.

Gem Stapleton, Peter Rodgers, and John Howse (2011a). A general method for drawing area-proportional Euler diagrams. *Journal of Visual Languages & Computing*, 22(6) : 426-42.

Gem Stapleton, Peter Rodgers, John Howse, and Leishi Zhang (2011b). Inductively Generating Euler Diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 17(1) : 88-100.

Gem Stapleton, Leishi Zhang, John Howse, and Peter Rodgers (2011c). Drawing Euler diagrams with circles: The theory of piercings. *IEEE Transactions on Visualization and Computer Graphics*, 17(7) : 1020-32.

Gem Stapleton, Jean Flower, Peter Rodgers, and John Howse (2012). Automatically drawing Euler diagrams with circles. *Journal of Visual Languages & Computing*, 23(3) : 163-93.

Margaret Mary Stark and Barry A Hollander (1990). Information graphics: Do they help readers understand news events? In: *Proceedings of the 73rd Annual Meeting of the Association for Education in Journalism and Mass Communication*, Minneapolis, MN, USA. ERIC.

Keith Stenning and Jon Oberlander (1995). A cognitive theory of graphical and linguistic reasoning: Logic and implementation. *Cognitive science*, 19(1) : 97-140.

John D Sterman (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2-3) : 291-330.

John D Sterman (2000a). Stocks and Flows. *Business dynamics: systems thinking and modeling for a complex world,* Irwin/McGraw-Hill, Boston, MA, USA, pp. 191-230.

John D Sterman (2000b). Dynamics of Stocks and Flows. *Business dynamics: systems thinking and modeling for a complex world,* Irwin/McGraw-Hill, Boston, MA, USA, pp. 231-62.

John D Sterman (2000c). Automobile Leasing Strategy: Gone Today, Here Tomorrow. *Business dynamics: systems thinking and modeling for a complex world,* Irwin/McGraw-Hill, Boston, MA, USA, pp. 42-54.

John D Sterman (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review*, 18(4) : 501-31.

John D Sterman and Linda Booth Sweeney (2007). Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change*, 80(3-4) : 213-38.

John D Sterman (2008). Risk communication on climate: mental models and mass balance. *Science*, 322(5901) : 532-33.

John D Sterman (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review*, 26(4) : 316-34.

John D Sterman (2011). Communicating climate change risks in a skeptical world. *Climatic Change*, 108(4) : 811-26.

Stanley S Stevens (1957). On the psychophysical law. *Psychological Review*, 64(3) : 153-81.

Stanley Smith Stevens (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects, 1st Edition*, John Wiley & Sons, New York, NY, USA.

Stanley Smith Stevens (1986). *Psychophysics: Introduction to its perceptual, neural, and social prospects, 2nd Edition*, Transaction Publishers, New Brunswick, NJ, USA.

James Stewart (2009). *Calculus, 6th Edition*, Thomas Brooks/Cole, Canada.

Gerald James Stine (1998). *Acquired Immune Deficiency Syndrome: Biological, Medical, Social, and Legal Issues, 3rd Edition*, Prentice Hall, Upper Saddle River, NJ, USA.

Eric R Stone, J Frank Yates, and Andrew M Parker (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, 3(4) : 243-56.

Eric R Stone, Winston R Sieck, Benita E Bull, Yates Frank, J, Stephanie C Parks, and Carolyn J Rush (2003). Foreground: background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, 90(1) : 19-36.

Kevin Stone (2013). *Optical Illusions*, BrainBashers. Accessed April 24, 2013, http://www.brainbashers.com/opticalillusions.asp.

Robert H Storer, S David Wu, and Renzo Vaccari (1992). New search spaces for sequencing problems with application to job shop scheduling. *Management Science*, 38(10) : 1495-509.

Paul Stothard and David S Wishart (2005). Circular genome visualization and exploration using CGView. *Bioinformatics*, 21(4) : 537-39.

Victor J Strecher, Todd Greenwood, Catharine Wang, and Dana Dumont (1999). Interactive Multimedia and Risk Communication. *JNCI Monographs*, 1999(25) : 134-39.

Avneesh Sud, Danyel Fisher, and Huai-Ping Lee (2010). Fast Dynamic Voronoi Treemaps. In: *Proceedings of the 7th International Symposium on Voronoi Diagrams in Science and Engineering (ISVD)*, pp. 85-94, Quebec, QC, Canada. IEEE.

Kozo Sugiyama and Kazuo Misue (1995a). Graph Drawing by the Magnetic Spring Model. *Journal of Visual Languages and Computing*, 6(3) : 217-31.

Kozo Sugiyama and Kazuo Misue (1995b). A simple and unified method for drawing graphs: Magnetic-spring algorithm. In: *Proceedings of the DIMACS International Workshop on Graph Drawing (GD 1994), Lecture Notes in Computer Science 894*, pp. 364-75, Princeton, NJ, USA. Springer.

Kozo Sugiyama (2002). *Graph drawing and applications for software and knowledge engineers*, vol. 11, World Scientific Publishing Co, Singapore.

Shipeng Sun (2013). An Optimized Rubber-Sheet Algorithm for Continuous Area Cartograms. *The Professional Geographer*, 65(1) : 16-30.

James Surowiecki (2005). *The Wisdom of Crowds*, Anchor Books, New York, NY, USA (surowiecki2005wisdom).

Linda Booth Sweeney and John D Sterman (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review*, 16(4) : 249-86.

John Sweller, Paul Chandler, Paul Tierney, and Martin Cooper (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, 119(2) : 176-92.

John Sweller (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4) : 295-312.

John Sweller, Merrienboer Van, Jeroen J G, and Fred G W C Paas (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3) : 251-96.

Jonathan Swinton (2007). *Venn diagrams with the Vennerable package*. https://r-forge.r-project.org/projects/vennerable, CVS document Id: Venn.Rnw, v 1.24 2007/04/23 22:10:13 js229 Exp. Accessed February 28, 2013, ftp://ftp.heanet.ie/disk1/sourceforge/v/project/ve/vennerable/Vignette/0.1/Venn.pdf.

Jonathan Swinton (2009). *Venn diagrams in R with the Vennerable package*. https://r-forge.r-project.org/projects/vennerable, CVS document Id: Venn.Rnw, v 1.26 2007/10/16 10:59:52 js229 Exp. Accessed February 28, 2013, https://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/Vennerable/inst/doc/Venn.pdf?revision=2&root=vennerable&pathrev=22.

Syracuse Post Standard (1911), 'Speakers Give Sound Advice', *Syracuse Post Standard*, March 28, 1911, p 18.

E-G Talbi and Traian Muntean (1993). Hill-climbing, simulated annealing and genetic algorithms: a comparative study and application to the mapping problem. In: *Proceedings of the 26th Hawaii International Conference on System Sciences, vol. 2*, pp. 565-73, Kauai, HI, USA. IEEE.

Roberto Tamassia, Battista Di, Giuseppe, and Carlo Batini (1988). Automatic graph drawing and readability of diagrams. *IEEE Transactions on Systems, Man and Cybernetics*, 18(1) : 61-79.

Laura G Tateosian, Christopher G Healey, and James T Enns (2007). Engaging Viewers Through Nonphotorealistic Visualizations. In: *Proceedings of the 5th International Symposium on Non-Photorealistic Animation and Rendering (NPAR)*, pp. 93-102, San Diego, CA, USA. ACM.

Soon Tee Teoh and Ma Kwan-Liu (2002). RINGS: A Technique for Visualizing Large Hierarchies. In: *Proceedings of the 10th International Symposium on Graph Drawing (GD 2002), Lecture Notes in Computer Science 2528*, pp. 268-75, Irvine, CA, USA. Springer.

Jérôme Thièvre, Marie-Luce Viaud, and Anne Verroust-Blondet (2005). Using Euler diagrams in traditional library environments. *Electronic Notes in Theoretical Computer Science*, 134: 189-202.

Kellom Tomlinson (1735). *The Art of Dancing Explained by Reading and Figure, Book II*, London, UK.

David F Treagust, Reinders Duit, Paul Joslin, and Ivo Lindauer (1992). Science teachers' use of analogies: Observations from classroom practice. *International Journal of Science Education*, 14(4) : 413-22.

Anne Treisman (1985). Preattentive Processing in Vision. *Computer Vision, Graphics, and Image Processing*, 31(2) : 156-77.

Anne Treisman and Janet Souther (1985). Search asymmetry: a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114(3) : 285-310.

Anne Treisman and Stephen Gormican (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95(1) : 15.

Anne M Treisman and Garry Gelade (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1) : 97-136.

Jennifer Tsai (2012), *Interactive visualizations to improve Bayesian reasoning*, PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Elisabet Tubau (2008). Enhancing probabilistic reasoning: The role of causal graphs, statistical format and numerical skills. *Learning and Individual Differences*, 18(2) : 187-96.

Edward Rolf Tufte (1983). *The Visual Display of Quantitative Information, 1st Edition*, Graphics Press, Cheshire, CT, USA.

Edward Rolf Tufte (1990). *Envisioning Information*, Graphics Press, Cheshire, CT, USA.

Edward Rolf Tufte (1997a). *Visual Explanations: Images and Quantities, Evidence and Narrative*, Graphics Press, Cheshire, CT, USA.

Edward Rolf Tufte (1997b). *Visual and Statistical Thinking: Displays of Evidence for Decision Making*, Graphic Press, Cheshire, CT, USA (tufte1997visual).

Edward Rolf Tufte (2001). *The Visual Display of Quantitative Information, 2nd Edition*, Graphics Press, Cheshire, CT, USA.

Edward Rolf Tufte (2006). *Beautiful Evidence*, Graphics Press, Cheshire, CT, USA.

John Wilder Tukey (1977). *Exploratory Data Analysis*, vol. 231, Addison-Wesley, Reading, MA, USA.

Daniel Tunkelang (1999), *A numerical optimization approach to general graph drawing*, PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA.

William Thomas Tutte (1960). Convex representations of graphs. *Proceedings London Mathematical Society*, 10(38) : 304-20.

William Thomas Tutte (1963). How to draw a graph. *Proceedings of the London Mathematical Society*, 13(3) : 743-68.

Amos Tversky and Daniel Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157) : 1124-31.

Amos Tversky and Daniel Kahneman (1980). Causal schemas in judgments under uncertainty. In Martin Fishbein (ed.), *Progress in Social Psychology,* Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

Amos Tversky and Daniel Kahneman (1982). Evidential impact of base rates. In Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.), *Judgment under uncertainty: Heuristics and biases,* Cambridge University Press, Cambridge, UK, pp. 153-60.

Amos Tversky and Daniel Kahneman (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4) : 293-315.

Matej Urbas and Mateja Jamnik (2011). Heterogeneous Proofs: Spider Diagrams Meet Higher-Order Provers. In: *Proceedings of the 2nd International Conference on Interactive Theorem Proving (ITP), Lecture Notes in Computer Science 6898*, pp. 376-82, Berg en Dal, The Netherlands. Springer.

Matej Urbas and Mateja Jamnik (2012). Diabelli: A Heterogeneous Proof System. In: *Proceedings of the 6th International Joint Conference on Automated Reasoning (IJCAR), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 7364*, pp. 559-66, Manchester, UK. Springer.

Matej Urbas, Mateja Jamnik, Gem Stapleton, and Jean Flower (2012). Speedith: a diagrammatic reasoner for spider diagrams. In: *Proceedings of the 7th International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 7352*, pp. 163-77, Canterbury, UK. Springer.

J Utech, Jürgen Branke, Hartmut Schmeck, and Peter Eades (1998). An evolutionary algorithm for drawing directed graphs. In: *Proceedings of the 2nd International Conference on Imaging Science, Systems and Technology (CISST)*, pp. 154-60, Las Vegas, NV, USA.

Marc van Kreveld and Bettina Speckmann (2007). On rectangular cartograms. *Computational Geometry*, 37(3) : 175-87.

Michael Florent van Langren (1644). *La Verdadera Longitud por Mar y Tierra*, Antwerp.

Jarke J van Wijk and Huub van de Wetering (1999). Cushion Treemaps: Visualization of Hierarchical Information. In: *Proceedings of the 5th IEEE Symposium on Information Visualization (InfoVis)*, pp. 73-78, San Francisco, CA, USA. IEEE.

Andrew Vande Moere and Helen Purchase (2011). On the role of design in information visualization. *Information Visualization - Special Issue on State of the Field and New Research Directions*, 10(4) : 356-71.

John Venn (1880). On the diagrammatic and mechanical representation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(59) : 1-18.

Tom Verguts, Wim Fias, and Michaël Stevens (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, 12(1) : 66-80.

Anne Verroust and Marie-Luce Viaud (2004). Ensuring the drawability of extended Euler diagrams for up to 8 sets. In: *Proceedings of the 3rd International Conference on the Diagrammatic Representation and Inference (Diagrams), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) 2980*, pp. 128-41, Cambridge, UK. Springer.

Bret Victor (2011). *Explorable Explanations*. Accessed May 4, 2013, http://worrydream.com/ ExplorableExplanations.

W Kip Viscusi (1992). *Smoking: Making the risky decision*, Oxford University Press, New York, NY, USA.

Vivianne H M Visschers, Ree M Meertens, Wim W F Passchier, and Nanne N K de Vries (2009). Probability Information in Risk Communication: A Review of the Research Literature. *Risk Analysis*, 29(2) : 267-87.

Chris Walshaw (2001). A multilevel algorithm for force-directed graph drawing. In: *Proceedings of the 8th International Symposium on Graph Drawing (GD 2000), Lecture Notes in Computer Science 1984*, pp. 171-82, Colonial Williamsburg, VA, USA. Springer.

Yuanyuan Y Wang, Rui B Chang, and Emily R Liman (2010). TRPA1 is a component of the nociceptive response to CO_2. *The Journal of Neuroscience*, 30(39) : 12958-63.

Colin Ware, Helen Purchase, Linda Colpoys, and Matthew McGill (2002). Cognitive measurements of graph aesthetics. *Information Visualization*, 1(2) : 103-10.

Colin Ware (2008). *Visual Thinking: for Design*, Morgan Kaufmann, San Francisco, CA, USA.

Colin Ware (2012). *Information Visualization: Perception for Design, 3rd Edition*, Morgan Kaufmann, Waltham, MA, USA.

Erika A Waters, N D Weinstein, G A Colditz, and Karen M Emmons (2007). Reducing aversion to side effects in preventive medical treatment decisions. *Journal of Experimental Psychology: Applied*, 13(1) : 11-21.

H Gilbert Welch and William C Black (2010). Overdiagnosis in Cancer. *Journal of the National Cancer Institute*, 102(9) : 605-13.

Charles M Wharton, Keith J Holyoak, Paul E Downing, Trent E Lange, Thomas D Wickens, and Eric R Melz (1994). Below the surface: Analogical similarity and retrieval competition in reminding. *Cognitive Psychology*, 26: 64-64.

B Wieland, M Wittwer, G Regula, TM Wassenaar, AP Burnens, J Keller, and KDC St{\"a}rk (2005). Phenon cluster analysis as a method to investigate epidemiological relatedness between sources of Campylobacter jejuni. *Journal of Applied Microbiology*, 100(2) : 316-24.

Wikimedia Foundation (2011). *File: British Isles Euler diagram 15.svg*, Wikipedia. Accessed April 1, 2013, http://en.wikipedia.org/wiki/File:British_Isles_Euler_diagram_15.svg (created by TWCarlson).

Wikimedia Foundation (2013). *Template: Supranational European Bodies*, Wikipedia. Accessed April 1, 2013, http://en.wikipedia.org/wiki/Template:Supranational_European_Bodies.

Leland Wilkinson (2012). Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Transactions on Visualization and Computer Graphics*, 18(2) : 321-31.

Gertrude Williams (1945). *Women and Work*, Nicholson & Watson.

Edward O Wilson (1998). *Consilience: The Unity of Knowledge*, Vintage Books, Random House Digital, Inc, New York, NY, USA.

Alexandra Z Worden, Jea-Hyeok Lee, Thomas Mock, Pierre Rouzé, Melinda P Simmons, Andrea L Aerts, Andrew E Allen, Marie L Cuvelier, Evelyne Derelle, Meredith V Everett, Elodie Foulon, Jane Grimwood, Heidrun Gundlach, Bernard Henrissat, Carolyn Napoli, Sarah M McDonald, Micaela S Parker, Stephane Rombauts, Aasf Salamov, Peter Von Dassow, Johathan H Badger, Pedro M Coutinho, Elif Demir, Inna Dubchak, Chelle Gentemann, Wenche Eikrem, Jill E Gready, Uwe John, William Lanier, Erika A Lindquist, Susan Lucas, Klaus F X Mayer, Herve Moreau, Fabrice Not, Robert Otillar, Olivier Panaud, Jasmyn Pangilinan, Ian Paulsen, Benoit Piegu, Aaron Poliakov, Steven Robbens, Jeremy Schmutz, Eve Toulza, Tania Wyss, Alexander Zelensky, Kemin Zhou, E Virginia Armbrust, Debashish Bhattacharya, Ursula W Goodenough, Yves Van de Peer, and Igor V Grigoriev (2009). Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes Micromonas. *Science*, 324(5924) : 268-72.

David F Wyatt, David Wynn, and John Clarkson (2009). Exploring Spaces of System Architectures Using Constraint-Based Classification and Euler Diagrams. In: *Proceedings of the 11th International Design Structure Matrix Conference (DSM)*, pp. 141-44, Geenville, SC, USA.

David F Wyatt (2010). *Set Visualiser*. Engineering Design Centre, University of Cambridge, Cambridge, UK. Accessed March 1, 2013, http://www-edc.eng.cam.ac.uk/tools/set_visualiser.

Kimihiko Yamagishi (2003). Facilitating Normative Judgments of Conditional Probability: Frequency or Nested Sets? *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, 50(2) : 97-106.

Chee K Yap (2006). Complete subdivision algorithms, I: Intersection of Bezier curves. In: *Proceedings of the 22nd Annual Symposium on Computational Geometry (SoCG)*, pp. 217-26, Sedona, AZ, USA. ACM.

Per-Henrik Zahl and Jan Maehlen (2004). Overdiagnosis in mammography screening. *Tidsskr Nor Lægeforen*, 124(17) : 2238-39.

Liye Zhang, Hong Ma, and B Franklin Pugh (2011). Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Research*, 21(6) : 875-84.

Caroline Ziemkiewicz and Robert Kosara (2008). The Shaping of Information by Visual Metaphors. *IEEE Transactions on Visualization and Computer Graphics*, 14(6) : 1269-76.

Caroline Ziemkiewicz and Robert Kosara (2009). Preconceptions and Individual Differences in Understanding Visual Metaphors. *Computer Graphics Forum*, 28(3) : 911-18.

Caroline Ziemkiewicz (2010), *Understanding the structure of information visualization through visual metaphors*, PhD thesis, University of North Carolina at Charlotte, Charlotte, NC, USA.

Caroline Ziemkiewicz and Robert Kosara (2010a). Beyond Bertin: Seeing the Forest despite the Trees. *IEEE Computer Graphics and Applications*, 30(5) : 7-11.

Caroline Ziemkiewicz and Robert Kosara (2010b). Implied Dynamics in Information Visualization. In: *Proceedings of the International Conference on Advanced Visual Interfaces (AVI)*, pp. 215-22, Rome, Italy. ACM.

Brian J Zikmund-Fisher, Brianna Sarr, Angela Fagerlin, and Peter A Ubel (2006). A Matter of Perspective: Choosing for Others Differs from Choosing for Yourself in Making Treatment Decisions. *Journal of General Internal Medicine*, 21(6) : 618-22.

Brian J Zikmund-Fisher, Angela Fagerlin, and Peter A Ubel (2008a). Improving Understanding of Adjuvant Therapy Options by Using Simpler Risk Graphics. *Cancer*, 113(12) : 3382-90.

Brian J Zikmund-Fisher, Peter A Ubel, Dylan M Smith, Holly A Derry, Jennifer B McClure, Azadeh Stark, Rosemarie K Pitsch, and Angela Fagerlin (2008b). Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient Education and Counseling*, 73(2) : 209-14.

Torre Dana Zuk (2008), *Visualizing Uncertainty*, PhD thesis, University of Calgary, Calgary, AB, Canada.

Appendix A

Experimental Results from Our Evaluation of Possible Cost Functions for euler*APE*

This appendix presents the experimental results from our evaluation in Section 5.6.4 subsection 'Evaluation of Possible Functions', based on which the cost function employed by our drawing method euler*APE was chosen. Further details and discussions on this evaluation are provided in Section 5.6.4 subsection 'Evaluation of Possible Functions'. As discussed in* Chapter 5, euler*APE, generates area-proportional 3-Venn diagrams drawn with ellipses.*

A.1 Evaluated Cost Functions

The following list of evaluated cost functions is the same as that in Section 5.6.4 subsection 'Evaluation of Possible Functions'. It has been provided in this appendix as a quick reference for the reader.

If

- ω is the area specification for which a diagram should be drawn,
- ω' is the area specification with the scaled quantities of ω obtained as defined in Section 5.6.2 indicating the required zone areas in the required good diagram for ω,
- d is a wellformed area-proportional diagram that is explored for ω during the optimization,
- Z is the set of labels describing the required set of zones interior to the curves of the diagram,
- $\omega'(z) \in \mathbb{R}^+$ is the area that $z \in Z$ should have in the required good diagram, and
- $A(z) \in \mathbb{R}^+$ is the area of $z \in Z$ in d computed using our analytic method M2 in Section 5.4.3,

then the evaluated cost functions were:

- F1. The statistical *stress* function used in `venneuler` [Wilkinson, 2012] that was computed with the source code of `venneuler` version 1.1-0 (provided to us by the author) but using $\omega'(z)$ and $A(z)$ for the required and the actual area of a zone $z \in Z$;

- F2. The first component of the 'idealistic' function of Chow and Rodgers [2005], based on *zoneError* Equation (5.17) Section 5.6.3, which we computed as

$$\sum_{z \in Z} \left(zoneError(z) \right)^2 \;=\; \sum_{z \in Z} \left(\frac{\omega(z)}{\omega_S} - \frac{A(z)}{A_S} \right)^2 \; where \; \omega_S = \sum_{z \in Z} \omega(z) \; and \; A_S = \sum_{z \in Z} A(z) \;;$$

- F3. The sum of the squared relative error of the zones, which we computed as

$$\sum_{z \in Z} \left(\frac{\omega'(z) - A(z)}{\omega'(z)} \right)^2 \;;$$

- F4. The maximum of the relative error of the zones, which we computed as

$$\max_{z \in Z} \left[\frac{|\omega'(z) - A(z)|}{\omega'(z)} \right] \;;$$

- F5. The sum of the relative error of the zones, which we computed as

$$\sum_{z \in Z} \frac{|\omega'(z) - A(z)|}{\omega'(z)} \;;$$

- F6. Our non-dimensionless cost function Equation (5.20) defined in Section 5.6.4 subsection 'Our Novel Functions' as

$$\frac{1}{|Z|} \sum_{z \in Z} \frac{\left(\omega'(z) - A(z) \right)^2}{A(z)} \;;$$

- F7. Our dimensionless cost function Equation (5.21) defined in Section 5.6.4 subsection 'Our Novel Functions' as

$$\frac{1}{|Z|} \sum_{z \in Z} \left(\frac{\omega'(z) - A(z)}{A(z)} \right)^2 \;;$$

- F8. Our dimensionless cost function Equation (5.22) defined in Section 5.6.4 subsection 'Our Novel Functions' as

$$\frac{1}{|Z|} \sum_{z \in Z} \frac{|\omega'(z) - A(z)|}{A(z)} \;.$$

A.2 Experimental Results

DL*as* and ASL*as* refer to the set of 10,000 area specifications obtained from respectively library DIAGLIB_*design* and library AREASPECLIB_*design* (details about the libraries in Section 5.5).

All the experiments were run on an Intel Core i7-3770 CPU @3.4GHz with 8GB RAM, 64-bit Microsoft Windows 7 Professional SP1 and Java Platform 1.7.0_10.

A.2.1 Accurate, Good Diagrams

Figure A.1 illustrates the number of accurate, good diagrams satisfying our diagram goodness measure Equation (5.19) in Section 5.6.3 that were generated for DL*as* and ASL*as* using cost functions F1-F8. So, the optimization terminated as soon as this diagram was obtained. The importance of squaring the inaccuracy of the zones in depicting the required quantity, such that large errors are heavily weighted, is evident, as F4, F5 and F8 all had a very low number of good diagrams compared to other functions, and they were the least effective in the generating a good diagram when one was known to exist (i.e., for DL*as*), suggesting that they often encountered a local minimum as inappropriate paths were not easily identified. It is also interesting that F4 performed marginally better than F5, even though F4 only considers the maximum cost of all the zones. However, the success rates of F4 and F5 are very low.
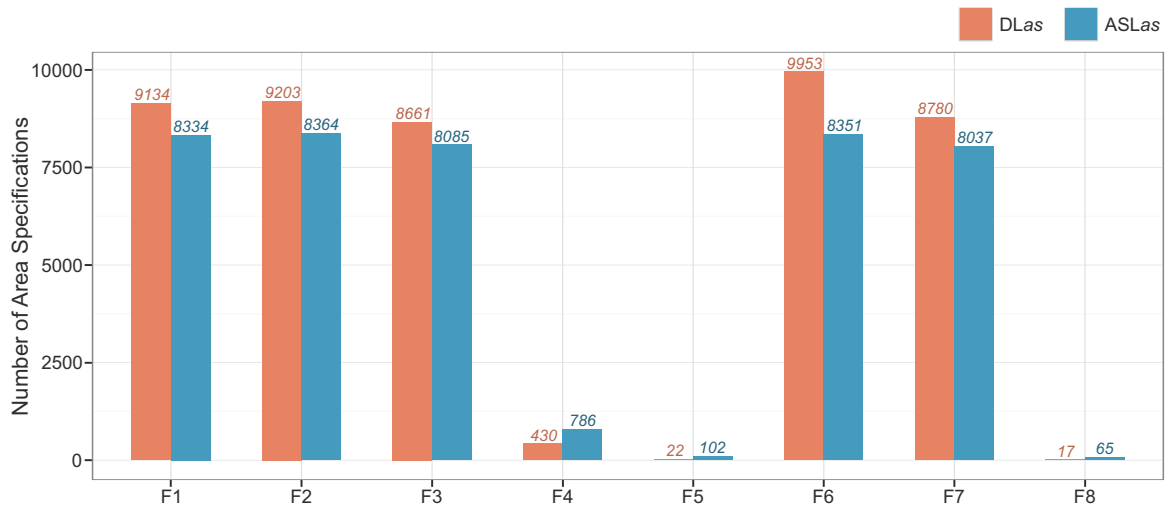
Figure A.1: *The number of area specifications in DLas and ASLas for which a good diagram (i.e., with diagError ≤ 10⁻⁶) was generated using the different cost functions F1-F8.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. These diagrams were good and so, they had a *diagError* $\leq 10^{-6}$. Thus, these diagrams satisfied our diagram goodness measure Equation (5.19) in Section 5.6.3 and their zone areas were accurately and directly proportional to the quantities in the required area specification.

These differences in the functions are also noticeable in Figure A.2, which illustrates the number of generated diagrams that had a *diagError* ≤ 0.01 (i.e., 1% as *diagError* is always in [0,1]). As defined by Equation (5.19) in Section 5.6.3, a good diagram must have a *diagError* $\leq 10^{-6}$ and so, Figure A.2 also includes the diagrams in Figure A.1, which we indicate with striped bars. As discussed in Section 5.6.5, the starting diagrams typically have a *diagError* ≤ 0.05 and so, analysing diagrams that after optimization have a *diagError* ≤ 0.01 is more adequate than analysing others with a greater *diagError*.
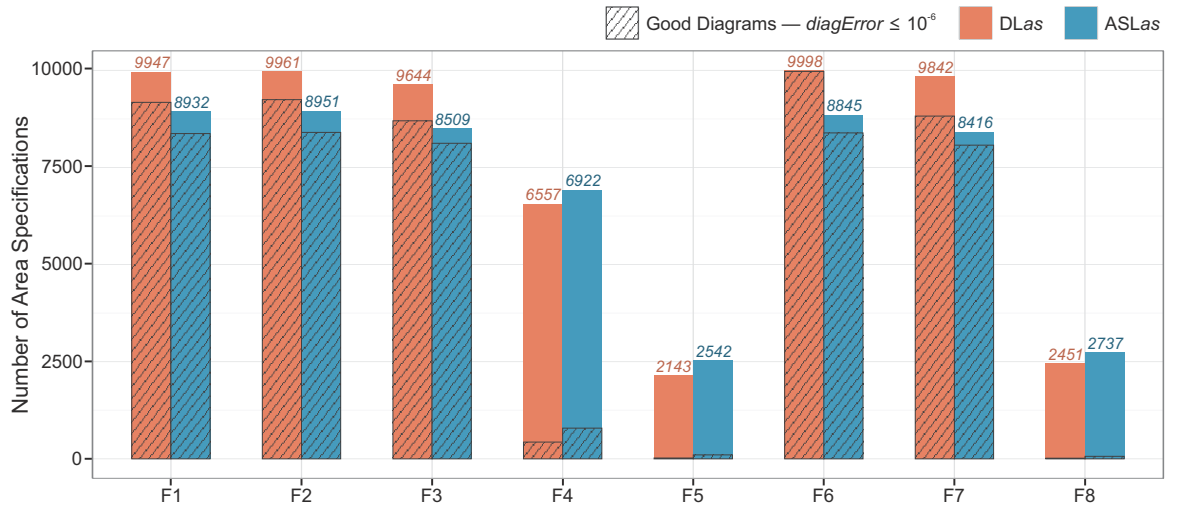
Figure A.2: *The number of area specifications in DLas and ASLas for which a diagram with diagError ≤ 0.01 was generated using the different cost functions F1-F8.* DLas and ASLas are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. Accurate, good diagrams with respect to an area specification satisfy our diagram goodness measure Equation (5.19) in Section 5.6.3 and thus, have a *diagError* ≤ 10^{-6}. So, this plot also includes the area specifications in Figure A.1 for which a good diagram was generated. These area specifications (i.e., for which a good diagram was generated) are indicated on this plot with striped bars.

Focusing on the results of the other cost functions, in Figure A.1 we note that for DL*as*, our non-dimensionless cost function F6 outperformed all the rest by generating good diagrams for 99.5% of the area specifications, followed by F2, F1, F7 and F3 which generated good diagrams for respectively 92.0%, 91.3%, 87.8% and 86.6% of the area specifications. The same is evident when all the generated diagrams with *diagError* ≤ 0.01 are considered, which, as shown in Figure A.2, F6, F2, F1, F7 and F3 generated such diagrams for respectively 99.8%, 99.6%, 99.5%, 98.4% and 96.4% of the area specifications. As for ASL*as*, the results for F2, F6 and F1 were close as they generated good diagrams for respectively 83.6%, 83.5% and 83.3% of the area specifications, followed by F3 and F7 with 80.9% and 80.4% respectively. Considering all the generated diagrams with *diagError* ≤ 0.01, Figure A.2 illustrates that F1 marginally surpassed F6, so that F2, F1, F6, F3 and F7 generated such diagrams for respectively 89.5%, 89.3%, 88.5%, 85.1% and 84.2% of the area specifications.

Hence, the cost functions that led to the highest percentages of good diagrams were F1, F2 and F6. The differences between these percentages are statistically significant for DL*as* but not for ASL*as* (using R package prop.test with Yates' continuity correction disabled, $\chi^2(2) = 767.8$, $p < 2.2 \times 10^{-16}$ for DL*as* and $\chi^2(2) = 0.3$, $p = 0.85$ for ASL*as*). The 99% confidence interval estimate of the difference between the 99.5% of F6 and 92.0% of F2 both for DL*as* is between 6.8% and 8.2% ($\chi^2(1) = 695.8$, $p < 2.2 \times 10^{-16}$) and similarly the 99% confidence interval estimate of the difference between the 99.5% of F6 and 91.3% of F1 both for DL*as* is between 7.4% and 8.9% ($\chi^2(1) = 769.8$, $p < 2.2 \times 10^{-16}$). It is also interesting that the difference between the percentages of F1 and F2 are not statistically significant for both DL*as* ($\chi^2(1) = 3.1$, $p = 0.08$) and ASL*as* ($\chi^2(1) = 0.3$, $p = 0.57$).

A.2.2 Effect of Cost Function on *diagError*

Cost Functions F1-F8

The optimization algorithm terminates as soon as a diagram satisfies our diagram goodness measure (Equation (5.19), Section 5.6.3) and so, all the good diagrams had a *diagError* $\leq 10^{-6}$. So, to better understand the effect of the cost function on the *diagError* of the generated diagrams, we should consider the generated diagrams that were non-good (i.e., with *diagError* > 10^{-6}). Boxplots of the log_{10}(*diagError*) of such diagrams per cost function are shown in Figure A.3 for DL*as* and ASL*as*.
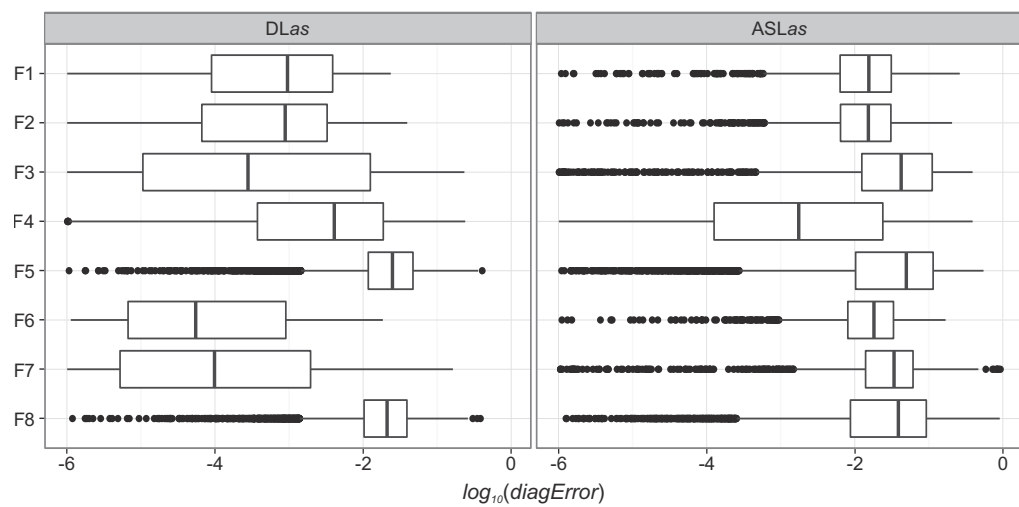
Figure A.3: *The log_{10}(diagError) of non-good diagrams generated for DLas and ASLas using the different cost functions F1-F8.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. These diagrams were non-good, so they had a *diagError* > 10^{-6}. The diagrams generated for DL*as* had the following *diagError* means: F1, 2.70×10^{-3} (*N* = 866); F2, 2.48×10^{-3} (*N* = 797); F3, 1.35×10^{-2} (*N* = 1339); F4, 1.52×10^{-2} (*N* = 9570); F5, 3.57×10^{-2} (*N* = 9978); F6, 1.77×10^{-3} (*N* = 47); F7, 4.72×10^{-3} (*N* = 1220); F8, 2.95×10^{-2} (*N* = 9983). The diagrams generated for ASL*as* had the following *diagError* means: F1, 2.86×10^{-2} (*N* = 1666); F2, 2.39×10^{-2} (*N* = 1636); F3, 7.03×10^{-2} (*N* = 1915); F4, 2.41×10^{-2} (*N* = 9214); F5, 7.05×10^{-2} (*N* = 9898); F6, 2.36×10^{-2} (*N* = 1649); F7, 4.95×10^{-2} (*N* = 1963); F8, 5.79×10^{-2} (*N* = 9935).

For DL*as*, Figure A.3 illustrates that our cost functions F6 and F7 had the lowest medians. As expected, the worst medians were those of F4, F5 and F8, which were the least effective in generating good diagrams. Amongst the other five functions (i.e., F1, F2, F3, F6 and F7), we note that F1 and F2 had the highest medians and a similar interquartile range, while F3 had the largest maximum and the largest interquartile range. As for means, F6, F2 and F1 had the lowest means followed by F7 and F3 (*diagError* means respectively 1.77×10^{-3}, 2.48×10^{-3}, 2.70×10^{-3}, 4.72×10^{-3}, 1.35×10^{-2}). Besides having the lowest median (*diagError* = 5.49×10^{-5}) and mean (*diagError* = 1.77×10^{-3}), F6 had the lowest maximum (*diagError* = 1.85×10^{-2}) and its median was lower than the first quartile of F2, though F6 had the fewest non-good diagrams (*N* = 47).

For ASL*as*, Figure A.3 illustrates that the medians and the interquartile ranges of all the cost functions excluding F4 were similar, particularly for F1, F2 and F6 whose means were also very much alike (*diagError* medians respectively 1.55×10^{-2}, 1.53×10^{-2} and 1.82×10^{-2}; *diagError* means respectively 2.86×10^{-2}, 2.39×10^{-2} and 2.36×10^{-2}). The errors of all the other cost functions, excluding F4, F5 and F8, were much greater than those of DL*as*. This is so, as while a good diagram can be drawn for all the area specifications in DL*as* (since the area specifications are obtained from wellformed 3-Venn diagrams drawn with ellipses), there could be area specifications in ASL*as* for which a good diagram drawn with ellipses does not exist (since the area specifications are made up of random values). The errors of F4, F5 and F8 were however similar to those of DL*as*. This indicates the inadequacy of F4, F5 and F8 in directing the optimization to an adequate diagram, as their errors for area specifications for which a good diagram is known to be drawable (i.e., those in DL*as*) are as bad as those of area specifications for which a good diagram drawn with ellipses might not exist (i.e., those in ASL*as*).

Cost Functions F1, F2 and F6

Considering the effect of F1, F2 and F6 on the *diagError* of all the 10,000 diagrams generated for each of DL*as* and ASL*as* by each of the cost functions, a Friedman rank sum test for non-normal distributions and repeated-measure data revealed a significant effect of cost function on *diagError* for DL*as* ($\chi^2(2) = 2723.8$, $p < 2.2 \times 10^{-16}$) but not for ASL*as* ($\chi^2(2) = 4.1$, $p = 0.13$). A post-hoc test using Wilcoxon tests with Bonferroni correction showed significant differences with a medium effect size between F6 and F2 ($W = 10575265$, $Z = -49.83$, $p < 2.2 \times 10^{-16}$, $r = 0.35$) and between F6 and F1 ($W = 11127260$, $Z = -47.93$, $p < 2.2 \times 10^{-16}$, $r = 0.34$), but with a very small effect size between F2 and F1 ($W = 24678752$, $Z = 2.53$, $p = 0.035$, $r = 0.05$). If Pearson's correlation coefficient r is 0.3, it is a medium effect that explains 9% of the variance, and if r is 0.1, it is a small effect that explains 1% of the variance ($r = 0.5$ is a large effect) [Cohen, 1988, 1992; Field, 2009]. So, F6 indeed outperformed F1 and F2 in generating diagrams with a lower *diagError*, while F1 and F2 are indeed similar.

Overall Indications

All of these results, indicate that our non-dimensionless cost function F6 is the most effective in avoiding and handling local minima such that a good diagram is obtained when one is know to exist, but it is unclear which between F1, F2 and F6 is the most effective when it is unknown whether a good diagram for the given area specification exists. Also, though different, F1 and F2 seem to be similar in effectiveness. F3 is less effective than F1, F2 and F6 and marginally worse than F7 (though differences in *diagError* of the 10,000 diagrams generated by F3 and F7 were not statistically significant; Friedman test, $\chi^2(1) = 0.3$, $p = 0.58$ for DL_*as*, $\chi^2(1) = 2.6$, $p = 0.10$ for ASL*as*).

A.2.3 Avoiding Local Minima

In an informal experimentation, we had observed that previously proposed cost functions are inappropriate in identifying and avoiding paths that reduce the overall error of the diagram at the expense of constantly diminishing one zone that should be larger, up to the point when the zone becomes close to non-existent and the optimization terminates as a local minimum is reached. Thus, we devised our cost functions F6, F7 and F8 to be able to handle such cases effectively (as discussed in Section 5.6.4 subsection 'Our Novel Functions').

To determine whether our observation is true and whether our cost functions, particularly F6, are capable of avoiding such paths, we computed the smallest actual-to-required ratio of the zones, that is

$$min\,A/R \;\; = \;\; \min_{z \in Z}\left[\frac{A(z)}{\omega'(z)} \right] \tag{A.1}$$

where Z is the set of zone labels, $\omega'(z)$ is the required zone area and $A(z)$ is the actual zone area (as defined at beginning of this section), for every diagram generated by the different cost functions.

As shown in Figure A.4 and Figure A.5, $minA/R$ for a good, accurate diagram will be a value close to one and never close to zero (for the eight cost functions, median 1 and mean in [1,1.02] for DL*as* with $N = 46202$, median and mean both in [1,1.04] for ASL*as* with $N = 42125$). The diagram goodness measure (Equation (5.19), Section 5.6.3) takes into account the zone area relative to the total area of the diagram and so, a diagram could still be accurate and good when $minA/R$ is close to but not exactly one. A value less than one means that there is one or more zones that are smaller than required and a value greater than one means that all the zones are larger than required. So, if $minA/R$ is a value that is close to zero, then the diagram has a zone that is barely visible even though its required area is larger, indicating that the optimization was directed to a path that reduces the overall error of the diagram at the expense of this zone.
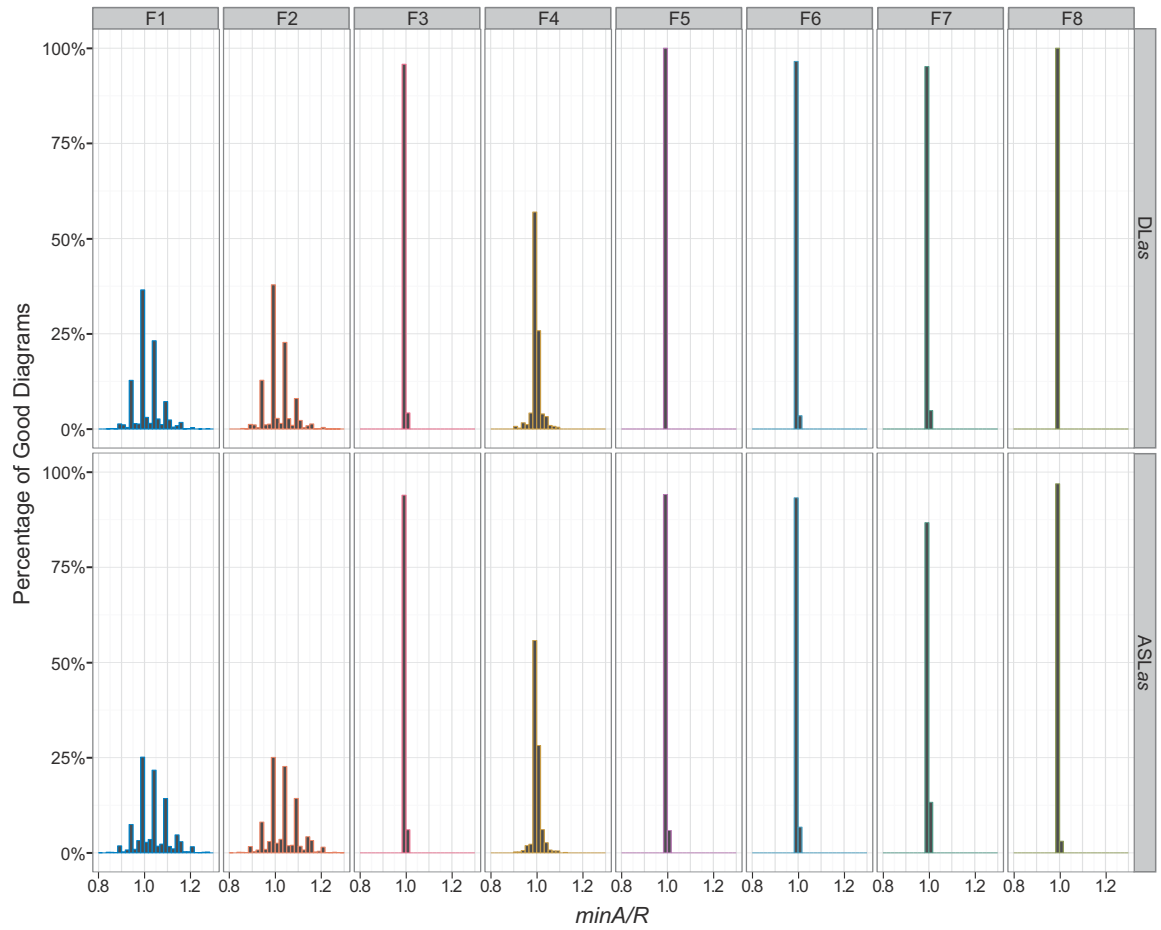
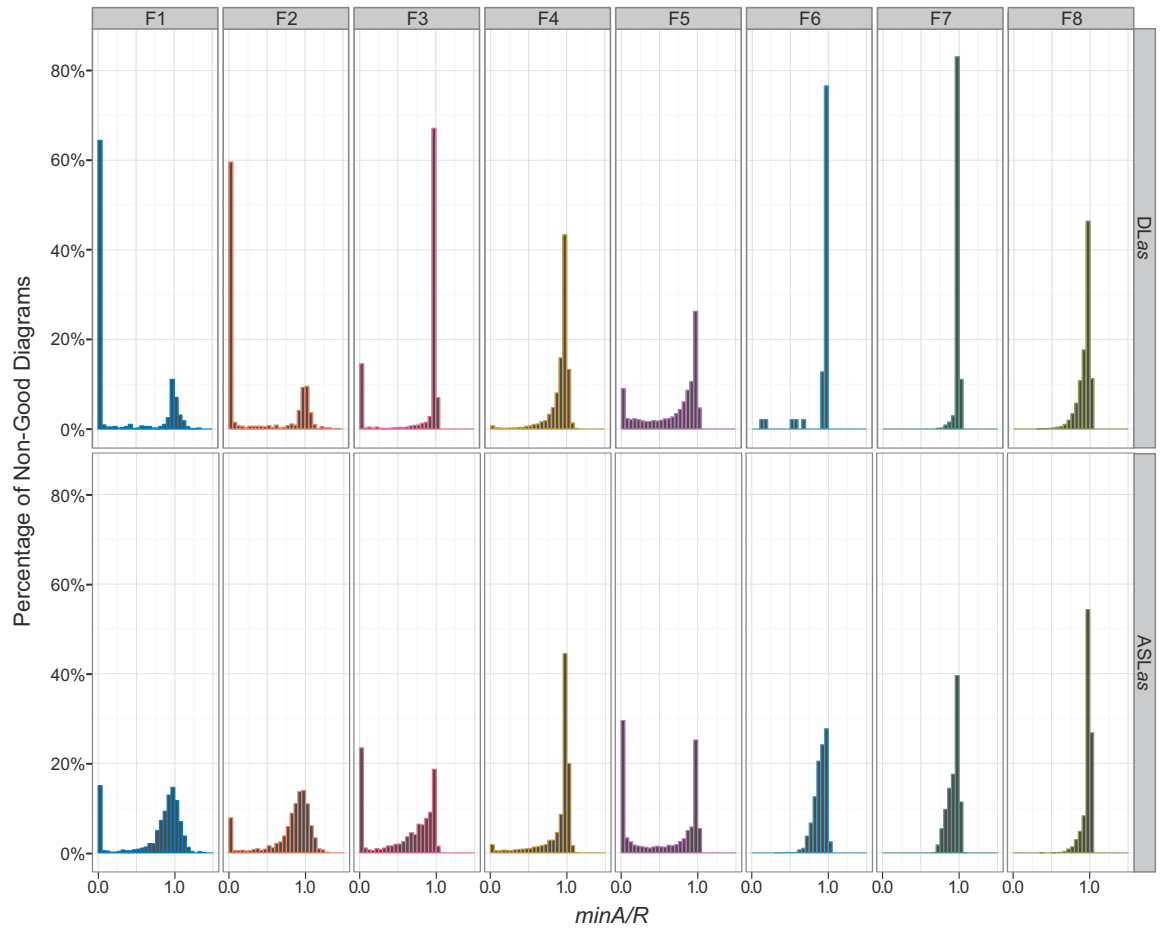**Figure A.4**: *The percentage distributions, relative to the number of good diagrams generated using each cost function F1-F8 for each of DLas and ASLas, of the smallest actual-to-required zone area ratio (minA/R) of the generated good diagrams.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. These diagrams were good, so they had a *diagError* ≤ 10^{-6}. The smallest actual-to-required zone area ratio (*minA/R*) was calculated using in Equation A.1. The total number of good diagrams generated by the cost functions is different: for DL*as*, F1 N = 9134, F2 N = 9203, F3 N = 8661, F4 N = 430, F5 N = 22, F6 N = 9953, F7 N = 8780, F8 N = 17; for AL*as*, F1 N = 8334, F2 N =8364, F3 N = 8085, F4 N = 786, F5 N = 102, F6 N = 8351, F7 N =8037, F8 N = 65.
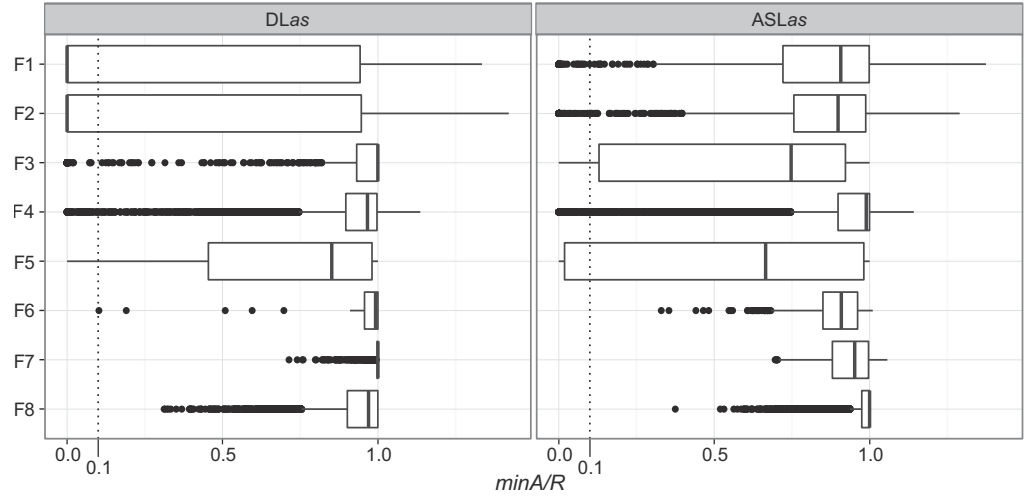
Figure A.5: *The smallest actual-to-required zone area ratio (minA/R) of the good diagrams generated for DLas and ASLas using the different cost functions F1-F8.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. These diagrams were good, so they had a *diagError* $\leq 10^{-6}$. The smallest actual-to-required zone area ratio (*minA/R*) was calculated using in Equation A.1. The total number of good diagrams generated by the cost functions is different: for DL*as*, F1 N = 9134, F2 N = 9203, F3 N = 8661, F4 N = 430, F5 N = 22, F6 N = 9953, F7 N = 8780, F8 N = 17; for AL*as*, F1 N = 8334, F2 N =8364, F3 N = 8085, F4 N = 786, F5 N = 102, F6 N = 8351, F7 N =8037, F8 N = 65.

The percentage distributions of *minA/R* for the generated non-good diagrams, in Figure A.6, illustrate that a considerable percentage of the non-good diagrams generated by all except for our cost functions F6, F7 and F8 had diagrams with *minA/R* close to zero (the first bar of the histograms in Figure A.6 represents the percentage of the non-good diagrams generated by the respective cost function that had a *minA/R* ≤ 0.05). This is particularly evident with the diagrams generated for DL*as*, for which a good diagram is known to exist. If a good diagram was not generated for such an area specification, then a local minimum was reached. Thus, the results obtained for DL*as* are more important for our purpose than those of ASL*as* as a good diagram might not exist for the latter.

The boxplots in Figure A.7 support the results in Figure A.6. The dotted line in the boxplots indicates that none of the diagrams generated by our cost functions F6, F7 and F8 had *minA/R* < 0.1 (the smallest *minA/R* was 0.1 for DL*as* only). All the other cost functions had at least one diagram with *minA/R* equal to zero, and between 1.0% to 65.4% of the generated non-good diagrams for DL*as* with *minA/R* < 0.1 and between 2.3% to 32.9% of the generated non-good diagrams for AL*as* with *minA/R* < 0.1 (see caption of Figure A.6 or Figure A.7 for the N values of each cost function for DL*as* and ASL*as*). F2 and F1, which had the second and the third largest set of good diagrams for DL*as* following F6, had the smallest mean and median for DL*as* (F2 mean = 0.35, median = 0, N = 797; F1 mean = 0.32, median = 0, N = 866). F5 and F3, both based on relative error, had the smallest median and mean for ASL*as* (F5 mean = 0.53, median = 0.67, N = 9898; F3 mean = 0.59, median = 0.75, N = 1915).

Figure A.6: *The percentage distributions, relative to the number of non-good diagrams generated using each cost function F1-F8 for each of DLas and ASLas, of the smallest actual-to-required zone area ratio (minA/R) of the generated non-good diagrams.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design,* each with 10,000 area specifications. These diagrams were non-good, so they had a *diagError* $> 10^{-6}$. The smallest actual-to-required zone area ratio (*minA/R*) was calculated using in Equation A.1. The total number of non-good diagrams generated by the cost functions is different: for DL*as*, F1 $N = 866$, F2 $N = 797$, F3 $N = 1339$, F4 $N = 9570$, F5 $N = 9978$, F6 $N = 47$, F7 $N = 1220$, F8 $N = 9983$; for AL*as*, F1 $N = 1666$, F2 $N = 1636$, F3 $N = 1915$, F4 $N = 9214$, F5 $N = 9898$, F6 $N = 1649$, F7 $N = 1963$, F8 $N = 9935$.

Figure A.7: *The smallest actual-to-required zone area ratio (minA/R) of the non-good diagrams generated for DLas and ASLas using the different cost functions F1-F8.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. These diagrams were good, so they had a *diagError* > 10^{-6}. The smallest actual-to-required zone area ratio (*minA/R*) was calculated using in Equation A.1. The total number of non-good diagrams generated by the cost functions is different: for DL*as*, F1 N = 866, F2 N = 797, F3 N = 1339, F4 N = 9570, F5 N = 9978, F6 N = 47, F7 N = 1220, F8 N = 9983; for AL*as*, F1 N = 1666, F2 N = 1636, F3 N = 1915, F4 N = 9214, F5 N = 9898, F6 N = 1649, F7 N = 1963, F8 N = 9935. The dotted line indicated that none of our cost functions F6, F7 and F8 have a *minA/R* less than 0.1.

Looking into the percentages of the non-good diagrams that had *minA/R* ≤ 10^{-6}, a value that is often considered close to zero (e.g., in eulerAPE and venneuler [Wilkinson, 2012]), we note that F1 and F2 had the greatest percentages for DL*as* followed by F3, F5 and F4 (36.0%, 30.0%, 12.2%, 2.0%, 0.1% respectively), while F3 and F1 had the greatest percentages for ASL*as* followed by F5, F2 and F4 (20.1%, 13.6%, 9.7%, 6.4%, 0.7% respectively). The total number of diagrams for each cost function and for each set of area specifications varied (see caption of Figure A.6 or Figure A.7 for the N values of each cost function for DL*as* and ASL*as*). Figure A.8 illustrates such non-good diagrams with *minA/R* ≤ 10^{-6} generated by F1, F2 and F3 (Figure A.8A-C) for ω = {*a*=17857, *b*=14021, *c*=35256, *ab*=898, *ac*=4080, *bc*=4, *abc*=1282} in DL*as* and the good diagram generated by our cost function F6 for ω (Figure A.8D). According to ω, zone *b* should have the smallest area, followed by zone *ab*, zone *abc* and then other zones. The zone areas of the diagrams generated by F1 and F2 (Figure A.8A and Figure A.8B respectively) are close to the required, except for the area of zone *bc* that is much larger than required and the area of zone *abc* that is very close to zero even though it is meant to be larger than that of zone *bc* and zone *ab*. The area of zone *abc* in the diagram generated by F3 (Figure A.8C) is also very close to zero. The other zone areas are not as close to the required areas as those of F1 and F2, with the exception of zone *bc* whose area is equal to that required. Due to the area of zone *abc*, *minA/R* was less than 10^{-6} for the three diagrams. F4, F5 and F8 also generated a non-good diagram for ω, but only the diagram of F5 had a *minA/R* ≤ 10^{-6} due to a very small zone *abc* with area close to zero. Same as F6, our cost function F7 generated a good diagram for ω. This illustrates the need for a cost function like ours that heavily weights zones whose area is very small with respect to that required.
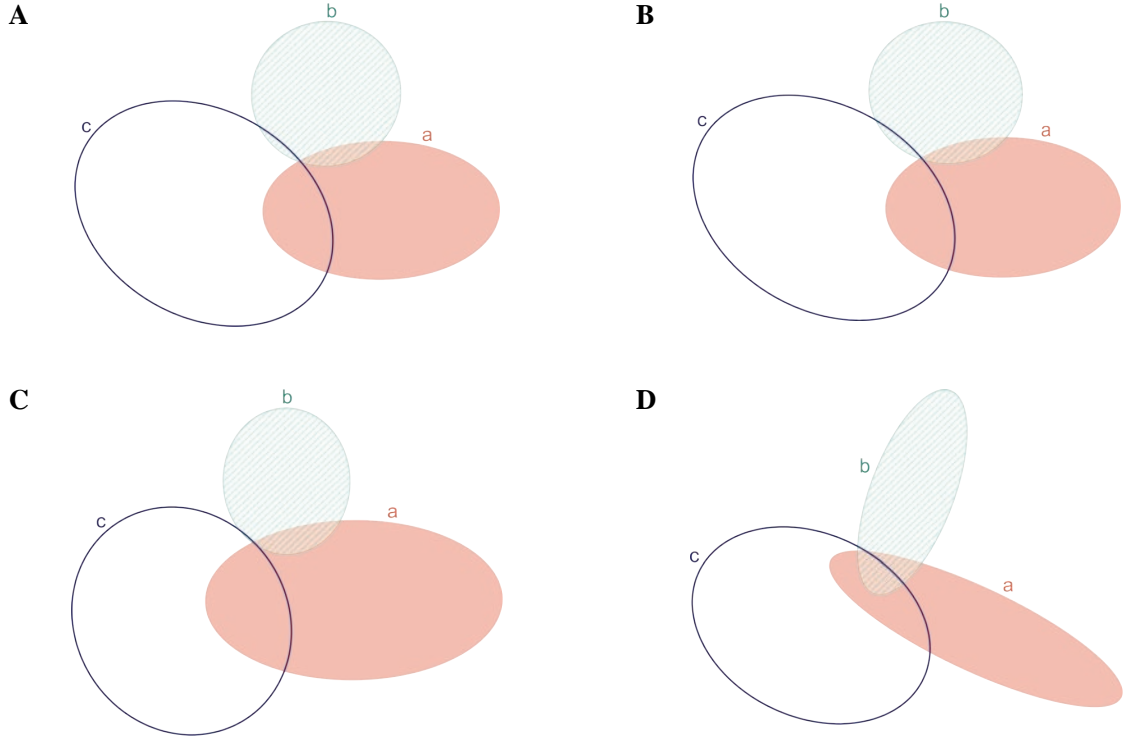
Figure A.8: *An example of an area specification in DLas for which a non-good diagram with the smallest actual-to-required zone area ratio (minA/R) less than or equal to 10^{-6} was generated using F1, F2 and F3, while a good diagram was generated using our cost function F6.* The diagrams were generated for area specification ω = {a = 17857, b = 14021, c = 35256, ab = 898, ac = 4080, bc = 4, abc = 1282} that was in DL*as*. Non-good diagrams were also generated for ω by F4, F5 and F8 (that of F5 had $minA/R$ = $3.3{\times}10^{-8}$), while a good diagram was generated by F7. (A) The non-good diagram generated by F1 for ω with zone areas {a = 16024, b = 12580, c = 31619, ab = 896, ac = 3695, bc = 209, abc = $1.30{\times}10^{-4}$} and with $minA/R$ = $3.93{\times}10^{-9}$ and $diagError$ = $1.75{\times}10^{-2}$. (B) The non-good diagram generated by F2 for ω with zone areas {a = 16613, b = 13080, c = 32624, ab = 1014, ac = 3945, bc = 250, abc = $1.36{\times}10^{-4}$} and with $minA/R$ = $4.12{\times}10^{-9}$ and $diagError$ = $1.75{\times}10^{-2}$. (C) The non-good diagram generated by F3 for ω with zone areas {a = 21707, b = 10135, c = 24077, ab = 1092, ac = 5706, bc = 4, abc = $5.82{\times}10^{-4}$} and with $minA/R$ = $1.76{\times}10^{-8}$ and $diagError$ = $1.03{\times}10^{-1}$. (D) The good diagram generated by F6 for ω with $minA/R$ = 1 and $diagError$ = $3.77{\times}10^{-7}$.

It is interesting that all the diagrams generated by F1, F2, F3 and F5 for ω had zone *abc* with an area very close to zero that led to a *minA/R* < 10^{-6}. In fact, considering the diagrams generated for DL*as*, for which area specifications a good diagram is known to exist, 64.6%, 68.1%, 50.9% and 42.6% of respectively the non-good diagrams generated by F1, F2, F3 and F5 with *minA/R* ≤ 10^{-6} had the actual-to-required area ratio for zone *abc* less than or equal to 10^{-6}. A possible reason why zone *abc* other than any other zone was diminished is that any change to the diagram during the optimization affects the area of the zone located in exactly the three ellipses and so, the smaller the required area of this zone with respect to the average required zone area of the diagram, the greater the likelihood that the overall error of the diagram is reduced at the expense of diminishing and increasing the error of this zone, as its contribution to the overall cost of the diagram would not be adequate to avoid such cases. As shown with our cost functions, such paths can easily be identified and avoided by heavily weighting those zones whose area is very small with respect to that required. To further aid the optimization in avoiding such paths, our starting diagram generator (explained in Section 5.6.5) minimizes the difference between the actual and the required area of the zone in exactly the three ellipses independently of the other zones. However, as seen with the results of this experiment, though this could help, it is not enough and a cost function like ours that identifies and avoids inappropriate paths is required.

These results indicate that most of the previously proposed cost functions, including `venneuler's` *stress* [Wilkinson, 2012] (F1) and Chow and Rodgers's [2005] 'idealistic' function (F2), as well as those based on relative errors (F3, F4 and F5), which are typically used in cartography but never for such diagrams, are all inadequate in identifying and avoiding paths that reduce the overall error of the diagram at the expense of heavily reducing the actual-to-required zone area ratio of one zone, up to the point when the zone is barely visible and the optimization terminates as it reaches a local minimum. These results also indicate that our cost function F6 and its dimensionless variants F7 and F8 are all capable of identifying and avoiding such paths.

A.2.4 Time and Iterations

Time is another important factor. It is also dependent on the number of iterations as the more iterations, the longer the convergence time. Figure A.9 and Figure A.10 respectively demonstrate log_{10}(time in seconds) and log_{10}(number of iterations) taken for good or non-good diagrams to be generated for DL*as* and ASL*as* using the different cost functions. Though the boxplots indicate that F4, F5 and F8 had a median that is among the lowest for both time and iterations, these cost functions generated very few good diagrams and are thus not effective. Considering the other functions, F3 and F7 had medians that were larger or close to the second quartile of F1, F2 and F6 for both time and iterations and considering that their percentages of good diagrams was also lower than those of the latter functions, then it is worth focusing only on F1, F2 and F6.
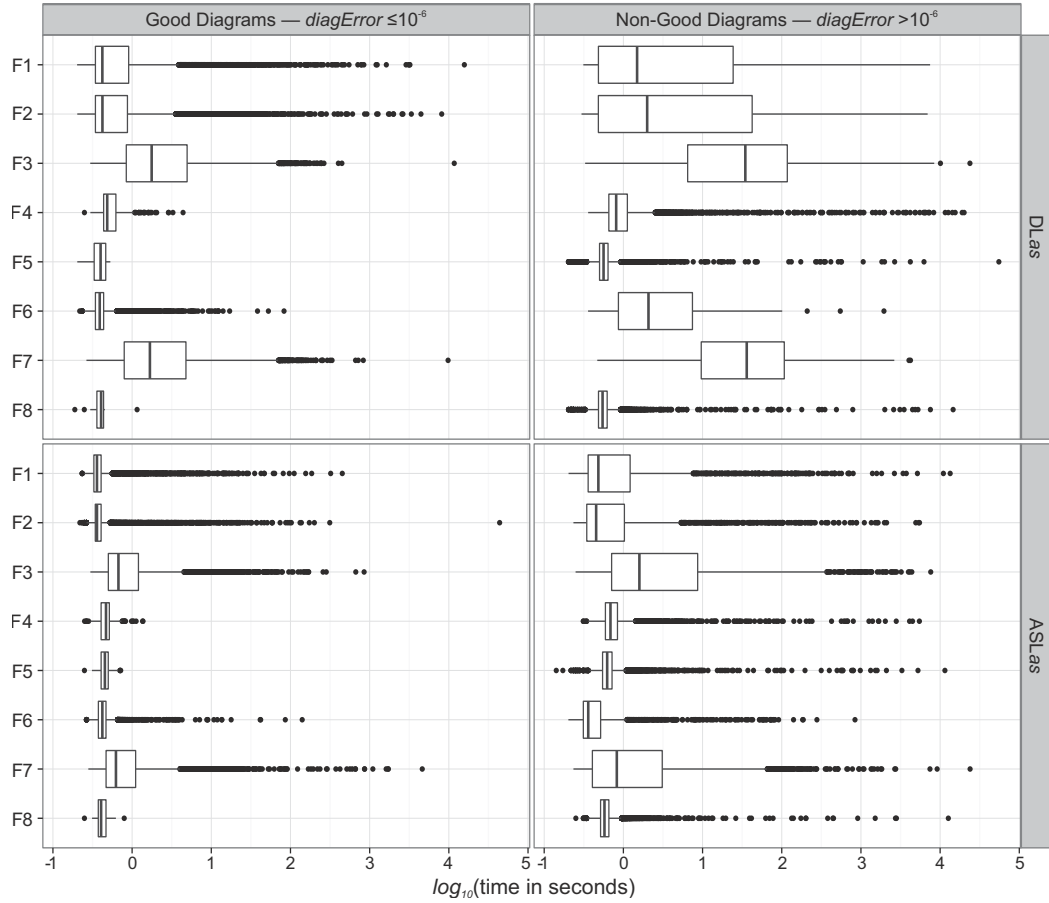
Figure A.9: *The log₁₀(time in seconds) for the generation of good and non-good diagrams for DLas and ASLas using the different cost functions F1-F8.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. Good diagrams have *diagError* ≤ 10^{-6}, while non-good diagrams have *diagError* > 10^{-6}. The total number of good diagrams generated by the cost functions is different: for DL*as*, F1 N = 9134, F2 N = 9203, F3 N = 8661, F4 N = 430, F5 N = 22, F6 N = 9953, F7 N = 8780, F8 N = 17; for AL*as*, F1 N = 8334, F2 N =8364, F3 N = 8085, F4 N = 786, F5 N = 102, F6 N = 8351, F7 N =8037, F8 N = 65. Similarly, the total number of non-good diagrams generated by the cost functions is different: for DL*as*, F1 N = 866, F2 N = 797, F3 N = 1339, F4 N = 9570, F5 N = 9978, F6 N = 47, F7 N = 1220, F8 N = 9983; for AL*as*, F1 N = 1666, F2 N = 1636, F3 N = 1915, F4 N = 9214, F5 N = 9898, F6 N = 1649, F7 N = 1963, F8 N = 9935.

Figure A.10: *The log_{10}(number of iterations) for the generation of good and non-good diagrams for DLas and ASLas using the different cost functions F1-F8.* DL*as* and ASL*as* are respectively the area specifications obtained from libraries DIAGLIB_*design* and AREASPECLIB_*design*, each with 10,000 area specifications. Good diagrams have *diagError* $\leq 10^{-6}$, while non-good diagrams have *diagError* $> 10^{-6}$. The total number of good diagrams generated by the cost functions is different: for DL*as*, F1 $N = 9134$, F2 $N = 9203$, F3 $N = 8661$, F4 $N = 430$, F5 $N = 22$, F6 $N = 9953$, F7 $N = 8780$, F8 $N = 17$; for AL*as*, F1 $N = 8334$, F2 $N = 8364$, F3 $N = 8085$, F4 $N = 786$, F5 $N = 102$, F6 $N = 8351$, F7 $N = 8037$, F8 $N = 65$. Similarly, the total number of non-good diagrams generated by the cost functions is different: for DL*as*, F1 $N = 866$, F2 $N = 797$, F3 $N = 1339$, F4 $N = 9570$, F5 $N = 9978$, F6 $N = 47$, F7 $N = 1220$, F8 $N = 9983$; for AL*as*, F1 $N = 1666$, F2 $N = 1636$, F3 $N = 1915$, F4 $N = 9214$, F5 $N = 9898$, F6 $N = 1649$, F7 $N = 1963$, F8 $N = 9935$.

From the boxplots of F1, F2 and F6 in Figure A.9 and Figure A.10, we note that most of the good diagrams were generated in less time and fewer iterations. F6 had the smallest maximum times compared to F1 and F2 (the maximum times of F6 were 12.4% and 24.0% of respectively those of F1 and F2 for DL*as* and 6.3% and 1.9% of respectively those of F1 and F2 for ASL*as*), while the three of them had the same minimum time (0.2 seconds). F6 also had the smallest maximum number of iterations compared to F1 and F2 (the maximums of F6 were 17.3% and 24.1% of respectively those of F1 and F2 for DL*as* and 11.7% and 3.1% of respectively those of F1 and F2 for ASL*as*), but not the smallest minimum number of iterations though the results were very much alike (the minimums for F1, F2 and F6 were 16, 20 and 20 for DL*as* and 21, 20 and 23 for ASL*as*).

Effect of Cost Function on Time

A Friedman rank sum test for non-normal distributions and repeated-measure data revealed a significant effect of F1, F2 and F6 on the generation time (in seconds) of the diagrams for DL*as* ($\chi^2(2) = 1574.7$, $p < 2.2 \times 10^{-16}$) and for ASL*as* ($\chi^2(2) = 1861.0$, $p < 2.2 \times 10^{-16}$). A post-hoc test using Wilcoxon tests with Bonferroni correction showed significant differences between every pair of cost functions.

For DL*as,* there was a medium effect size between F6 and F1 ($W = 10037550$, $Z = -48.64$, $p < 2.2 \times 10^{-16}$, $r = 0.34$) and between F6 and F2 ($W = 10202120$, $Z = -48.20$, $p < 2.2 \times 10^{-16}$, $r = 0.34$), and a very small effect size between F1 and F2 ($W = 24793228$, $Z = 5.95$, $p < 10^{-8}$, $r = 0.04$). For ASL*as,* there was a small effect size between F6 and F1 ($W = 29706974$, $Z = 22.76$, $p < 2.2 \times 10^{-16}$, $r = 0.16$) and between F6 and F2 ($W = 30735043$, $Z = 26.37$, $p < 2.2 \times 10^{-16}$, $r = 0.19$), and a very small effect size between F1 and F2 ($W = 22917786$, $Z = 4.56$, $p < 10^{-5}$, $r = 0.03$).

For DL*as*, the medians for F1, F2 and F6 were respectively 0.44 seconds, 0.44 seconds and 0.39 seconds, while their means were respectively 13.59 seconds, 15.23 seconds and 0.78 seconds. For ASL*as*, the medians for F1, F2 and F6 were respectively 0.37 seconds, 0.37 seconds and 0.41 seconds, while their means were respectively 7.20 seconds, 9.54 seconds and 0.85 seconds. So, though F6 had the lowest means, it did not have the lowest median for ASL*as*.

However, the effect size between F6 and both of F1 and F2 for ASL*as* was smaller ($r = 0.16$, $r = 0.19$) than that for DL*as* ($r = 0.34$, $r = 0.34$), and a small effect of 0.1 explains around 1% of the variance, while a medium effect of 0.3 explains around 9% of the variance [Cohen, 1988, 1992; Field, 2009]. Also, the difference between the means of F6 and those of F1 and F2 is greater than the difference between the medians. So, we can consider our cost function F6 to be more effective with respect to time than F1 and F2.

Effect of Cost Function on Iterations

Similar to time, a Friedman test revealed significant effect of F1, F2 and F6 on the number of iterations that were required for the generation of the diagrams for DL*as* ($\chi^2(2) = 1991.2$, $p < 2.2 \times 10^{-16}$) and ASL*as* ($\chi^2(2) = 1509.6$, $p < 2.2 \times 10^{-16}$). Significant differences between every pair of F1, F2 and F6 were reported by a post-hoc test using Wilcoxon tests with Bonferroni correction.

For DL*as,* there was a medium effect size between F6 and F1 ($W = 8414888$, $Z = -52.60$, $p < 2.2 \times 10^{-16}$, $r = 0.37$) and between F6 and F2 ($W = 8765361$, $Z = -51.27$, $p < 2.2 \times 10^{-16}$, $r = 0.36$), but a very small effect size between F1 and F2 ($W = 23814955$, $Z = 7.37$, $p < 10^{-12}$, $r = 0.05$). For

ASL*as* there was a small effect size between F6 and F1 ($W = 26598545$, $Z = 17.01$, $p < 2.2 \times 10^{-16}$, $r = 0.12$) and between F6 and F2 ($W = 27994146$, $Z = 21.54$, $p < 2.2 \times 10^{-16}$, $r = 0.15$), but a very small effect size between F1 and F2 ($W = 21569165$, $Z = 7.64$, $p < 10^{-13}$, $r = 0.05$).

For DL*as*, the medians for F1, F2 and F6 were 38, 37 and 32 respectively, while their means were 1245, 1431 and 73 respectively. For ASL*as*, the medians for F1, F2 and F6 were 33, 33 and 36 respectively, while their means were 1063, 1528 and 107 respectively.

These results are similar to those obtained for the generation time in that, F6 had much lower means than those of F1 and F2, and F6 had the lowest median for DL*as*, but a greater median for ASL*as*. However, once again the effect size between F6 and both of F1 and F2 for ASL*as* is smaller ($r = 0.12$, $r = 0.15$) than that for DL*as* ($r = 0.37$, $r = 0.36$), and the differences between the means are much greater than those between the medians. So, overall, our cost function F6 required fewer number of iterations to generate the diagrams.

Response Time Effectiveness

A response time of around 1 second ensures the users' train of thought is maintained and a response time of around 10 seconds ensures the users' attention is maintained [Card et al., 1991; Miller, 1968]. Considering the time taken to generate good diagrams for both DL*as* and ASL*as*, we find that F1, F2 and F6 respectively generated 85.8.7%, 86.1% and 98.3% of them in or less than 1 second and 97.7%, 97.8% and 99.9% of them in or less than 10 seconds ($N = 17468$, $N = 17567$, $N = 18304$ respectively). Considering the time taken to generate the non-good diagrams for both DL*as* and ASL*as*, we find that F1, F2 and F6 respectively generated 63.2%, 64.6% and 86.3% of them in or less than 1 second and 81.4%, 79.8% and 96.6% of them in or less than 10 seconds ($N = 2532$, $N = 2433$, $N = 1696$ respectively). So, our cost function F6 had the largest percentage of diagrams generated in or less than 1 second and in or less than 10 seconds, followed by F1 and F2 which has similar percentages. Thus, in this respect, F6 should be preferred as these results indicate that, using F6, in 97.3% of the cases (99% confidence interval 97.0% to 97.6%, $N = 20000$) a diagram is generated in or less than 1 second, and in 99.6% of the cases (99% confidence interval 99.5% to 99.7%, $N = 20000$) a diagram is generated in or less than 10 seconds.

F2 versus F6

As expected, overall, F2 required more time and iterations than F6 to generate a diagram for an area specification, particularly when the generated diagram was good (see Figure A.9 and Figure A.10). A possible reason for this is that, to calculate the error in the zones and to determine whether a change in the diagram will aid the optimization reach a good solution, F2 considers the zone area relative to the area of the diagram. As discussed in Section 5.6.4 subsection 'Our Novel Functions', this could be too restrictive as the area of most zones, especially at the initial stages of the optimization when major changes are explored in search for a good yet non-refined solution, will not be equal to that required and so, if the relative zone area is considered as in F2, other errors in the diagram will add up to the computed error of a zone. So, by considering the absolute difference between the required and the actual zone area independently of the other zones as in F6, convergence time and the number of the iterations could be reduced.

Appendix B

Materials for the Study Assessing the Effect of Euler Diagrams with Glyphs on Bayesian Reasoning

This appendix contains additional material to our user study on the effect of different Euler diagram visualizations on Bayesian reasoning, discussed in Chapter 6.

B.1 The Visualizations for the Bayesian Problems in Our Study

In our study, we compared a textual and six visual representations for three classic Bayesian problems using a diverse subject pool through crowdsourcing.

The visualizations included: Euler diagrams, glyphs, and combinations of both. The design of the visualizations (V1-V6) is discussed in Section 6.4.2. The visualizations were automatically and accurately generated using our diagram drawing software, eulerGlyphs, discussed in Section 6.3.

The three classic Bayesian problems included: the mammography problem [Eddy, 1982] (*Mam*), the cab problem [Tversky and Kahneman, 1982] (*Cab*), the choosing a course in economics problem [Ajzen, 1977] (*Eco*). The rationale of choosing these Bayesian problems is discussed in Section 6.4.3. The classic text that was used for these problems is provided in Table 6.1.

The following figures illustrate the visualizations (V1-V6) used for *Mam* (Figure B.1—replica of Figure 6.2 in Chapter 6), for *Cab* (Figure B.2) and for *Eco* (Figure B.3). These visualizations are also available at `http://www.aviz.fr/bayes`.
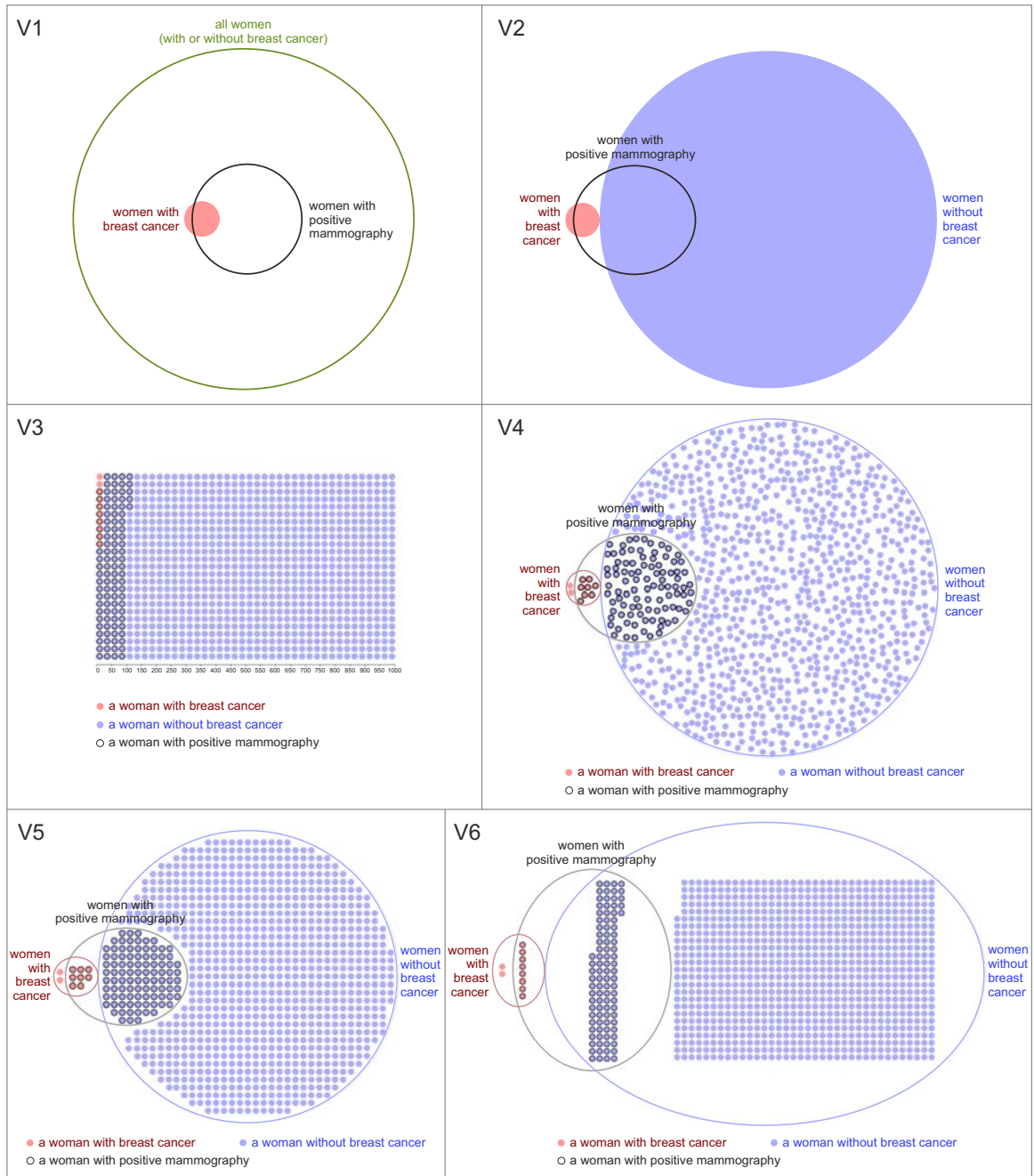
Figure B.1: *The six visualizations used for the Mam problem in our study.* The visualizations illustrate the classic mammography Bayesian problem [Eddy, 1982]. The diagrams were generated using our diagram drawing software, euler*Glyphs*. (V1) An area-proportional Euler diagram with the population of the problem represented as one set (depicted by the green curve). (V2) An area-proportional Euler diagram with the population of the problem represented as two sets (depicted by the red and blue curves). (V3) A frequency grid. (V4) An area-proportional Euler diagram with randomly positioned glyphs. (V5) A not area-proportional Euler diagram with uniformly positioned glyphs. (V6) A not area-proportional Euler diagram with frequency grid glyphs.

Figure B.2: *The six visualizations used for the Cab problem in our study.* The visualizations illustrate the classic mammography Bayesian problem [Tversky and Kahneman, 1982]. The diagrams were generated using our diagram drawing software, euler*Glyphs*. (V1) An area-proportional Euler diagram with the population of the problem represented as one set (depicted by the green curve). (V2) An area-proportional Euler diagram with the population of the problem represented as two sets (depicted by the red and blue curves). (V3) A frequency grid. (V4) An area-proportional Euler diagram with randomly positioned glyphs. (V5) A not area-proportional Euler diagram with uniformly positioned glyphs. (V6) A not area-proportional Euler diagram with frequency grid glyphs.
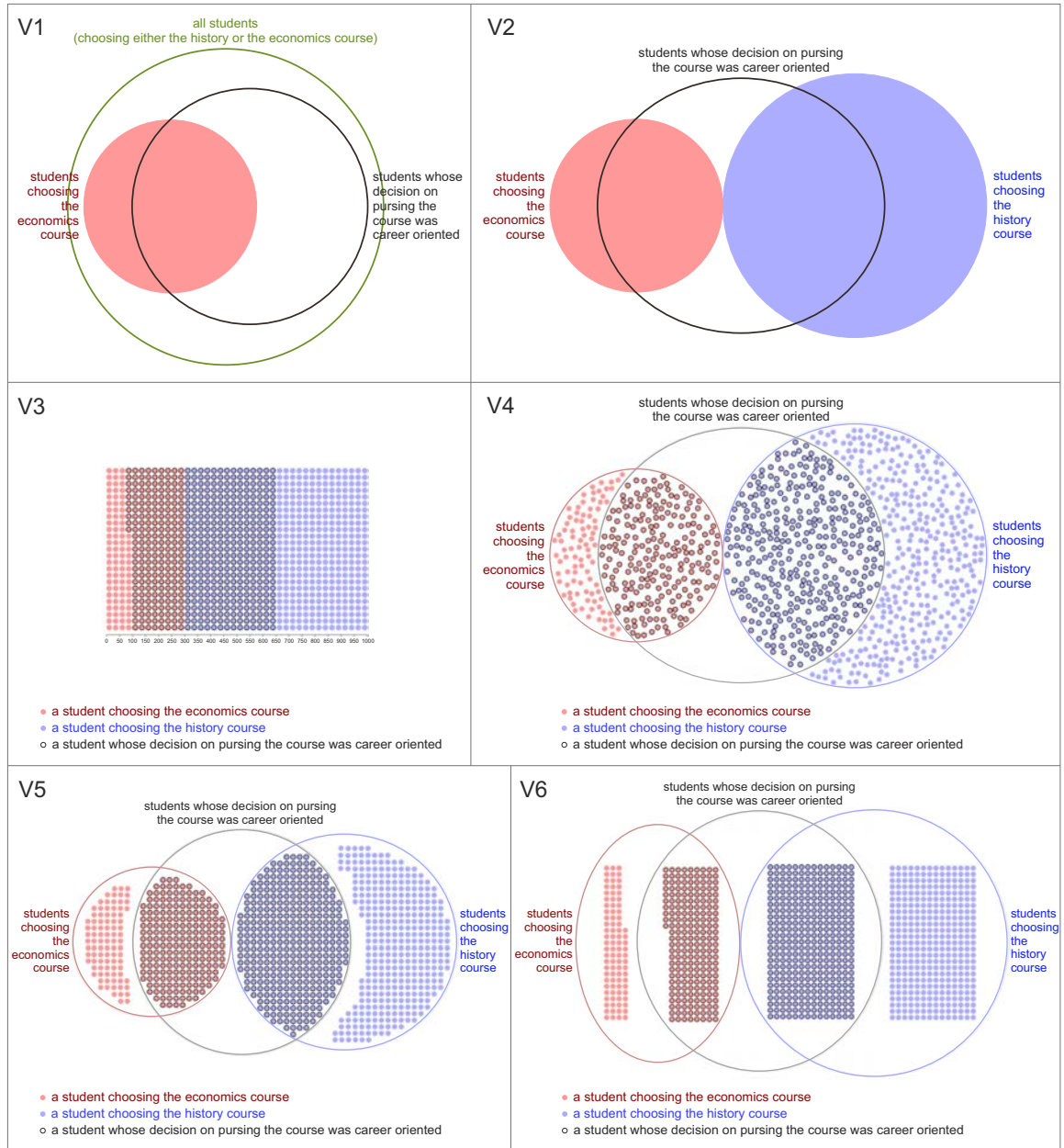
Figure B.3: *The six visualizations used for the Eco problem in our study.* The visualizations illustrate the classic mammography Bayesian problem [Ajzen, 1977]. The diagrams were generated using our diagram drawing software, euler*Glyphs*. (V1) An area-proportional Euler diagram with the population of the problem represented as one set (depicted by the green curve). (V2) An area-proportional Euler diagram with the population of the problem represented as two sets (depicted by the red and blue curves). (V3) A frequency grid. (V4) An area-proportional Euler diagram with randomly positioned glyphs. (V5) A not area-proportional Euler diagram with uniformly positioned glyphs. (V6) A not area-proportional Euler diagram with frequency grid glyphs.

B.2 The Tasks on Amazon Mechanical Turk in Our Study

The experiments were conducted through crowdsourcing using Amazon Mechanical Turk (MTurk) [Chen et al., 2011; Paolacci et al., 2010]. Thus, each task was implemented as a HIT (Human Intelligence Task).

The next sections illustrate an example of the HITs (including all the pages of the HIT) used in Experiment 1 (Appendix B.2.1; details in Section 6.5), qualitative feedback after Experiment 1 (Appendix B.2.2; details in Section 6.5.6), and Experiment 2 (Appendix B.2.3; details in Section 6.6). An online version of these HITs is also accessible through `http://www.aviz.fr/bayes`.

B.2.1 Experiment 1: Comparison of Visualizations

As discussed in Section 6.5.3, the HITs were made up of 10 pages, including: a page with instructions; three pages each with one of the three classic Bayesian problems we chose for our study, with no visualization or one of the visualizations V1-V6 discussed in Section 6.4.2; a page with catch questions; a page with the objective numeracy test; a page with the subjective numeracy test; two pages for the paper folding test; a final page with a brief questionnaire to learn the worker's demographics.

Each of the six problem orderings were tested with each of the seven visualization types and carried out by four different subjects. The following HIT illustrates problem ordering *Mam*, *Cab*, *Eco* and uses visualization type V4 (area-proportional Euler diagram with randomly positioned glyphs).

Figure B.4: *Page 1 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Figure B.5: *Page 2 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Problem 2

Please read the text below carefully and study the diagram, then answer the question as best as you can.

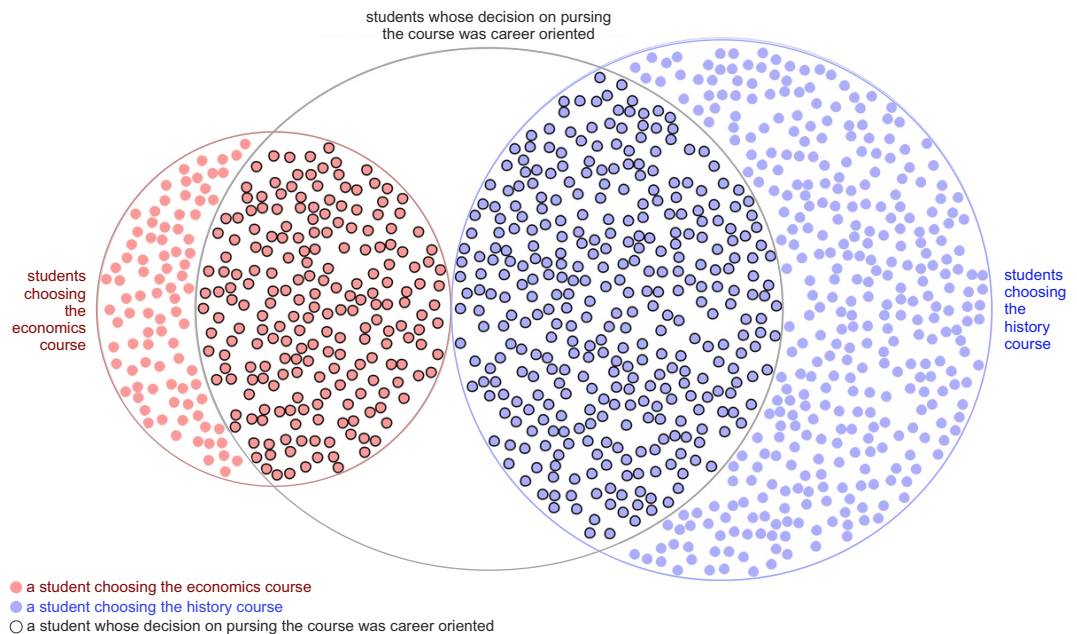Your answer does not need to be exact.

A cab was involved in a hit-and-run accident at night.
Two cab companies, the Red and the Blue, operate in the city.
Of every **100** cabs in the city, **15** are **Blue** and **85** are **Red**.

On the night of the accident, a witness identified the cab as **Blue**. The court tested the reliability of the witness under the similar visibility conditions with Blue and Red cabs.

When the cabs were **really Blue**, the witness **said they were Blue** in **12** out of **15** tests.
When the cabs were **really Red**, the witness **said they were Blue** in **17** out of **85** tests.

This is illustrated by the following diagram:
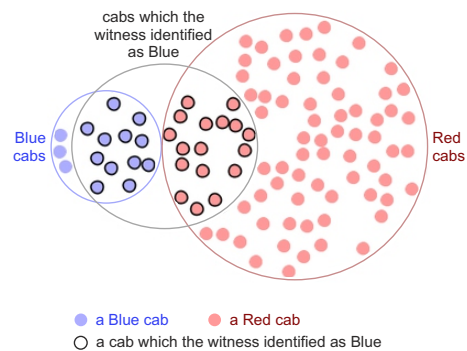
Please answer the following question:

What are the chances that the cab involved in the hit-and-run accident was **Blue**?

Your answer: [   ] out of [   ]

How confident are you in your answer?

No Confidence		Reasonable Confidence		Very High Confidence
○1	○2	○3	○4	○5

Next page (you cannot return to this page afterwards)

Figure B.6: *Page 3 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Problem 3

Please read the text below carefully and study the diagram, then answer the question as best as you can.

Your answer does not need to be exact.

In a small liberal arts college students take, as an elective, a general interest course in economics or history.

A recent analysis of enrollment figures showed that out of every **1,000** students,
700 students took the general interest course in **history**,
while **300** students took the course in **economics**.

For **210** of the students out of the **300** who took the **economics** course,
the decision on pursuing the economics course was **career oriented**.

For **350** of the students out of the **700** who took the **history** course,
the decision on pursuing the history course was **career oriented**.

This is illustrated by the following diagram:

students whose decision on pursing
the course was career oriented

students
choosing
the
economics
course

students
choosing
the
history
course

- a student choosing the economics course
- a student choosing the history course
- ○ a student whose decision on pursing the course was career oriented

Your answer: [   ] out of [   ]

How confident are you in your answer?

No Confidence		Reasonable Confidence		Very High Confidence
○1	○2	○3	○4	○5

Next page (you cannot return to this page afterwards)

Figure B.7: *Page 4 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Figure B.8: *Page 5 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Page 6/10

Smaller Problems

Please read the questions below and answer them as best as you can.

1. Imagine that you flip a coin 100 times. About how many times will the coin come up heads?

Your answer: ☐

2. After flipping a coin 10 times you have counted 7 heads and 3 tails. What is the chance (%) that your next flip will come up heads?

Your answer: ☐

3. 100 people have entered the Spring City Run. Seventy percent of the runners will finish the race. Of the 100 people who enter the race, how many will finish?

Your answer: ☐

4. In the Washington School raffle 5 people out of 100 who enter will win a prize. What percentage (%) of people who enter will win a prize?

Your answer: ☐

5. How much is one percent of 1000?

Your answer: ☐

6. One is what percent of 1000?

Your answer: ☐

Next page (you cannot return to this page afterwards)

Figure B.9: *Page 6 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Page 7/10

Which Presentation Do You Prefer?

For each of the following questions, please check the radio button that best reflects your answer.

1. When reading the newspaper, how **helpful** do you find tables and graphs that are parts of a story?

Not at all helpful					Extremely helpful
○1	○2	○3	○4	○5	○6

2. When people tell you the chance of something happening, do you prefer that they use **words** ("it rarely happens") or **numbers** ("there's a 1% chance")?

Always Prefer Words					Always Prefer Numbers
○1	○2	○3	○4	○5	○6

3. When you hear a weather forecast, do you prefer predictions using **percentages** (e.g., "there will be a 20% chance of rain today") or predictions using only **words** (e.g., "there is a small chance of rain today")?

Always Prefer Percentages					Always Prefer Words
○1	○2	○3	○4	○5	○6

4. How **often** do you find numerical information to be useful?

Never					Very Often
○1	○2	○3	○4	○5	○6

Next page   (you cannot return to this page afterwards)

Figure B.10: *Page 7 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Paper Folding Test

This is the last test. Please read the instructions below carefully before you proceed.

Instructions

In this exercise, you have to imagine the folding and unfolding of pieces of paper. Each problem has two figures on the left of a vertical line and others on the right of this line.

The **left figures** represent a square piece of paper being folded. The last of these figures has one or two small circles, to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point.

One of the five figures on the **right** of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct by choosing its corresponding letter.

The following is an example:
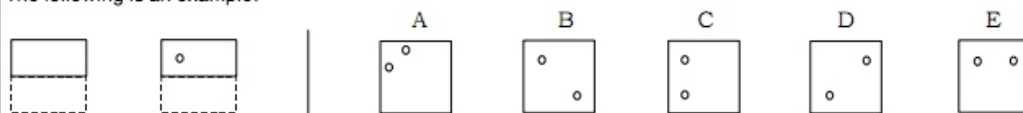
The correct answer is **C**. The folding, piercing and unfolding steps are shown below:
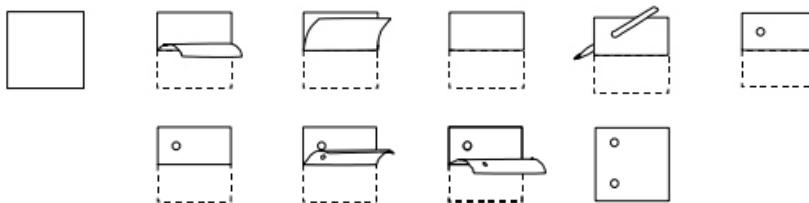
In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

You will have a **maximum of 3 minutes** to answer as many questions as you can. Click "Next Page" when you are ready.

Next page (you cannot return to this page afterwards)

Figure B.11: *Page 8 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Figure B.12: *Page 9 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

Page 10/10

Tell Us About You

Please note that we only know you by your worker ID and thus your anonymity will be protected at all times. Remember that you have to complete all the questions in the HIT to get your reward.

Gender: [▽]

Age: [ ]

Country of Residence: [ ]

Educational Level: [▽]

Occupation: [ ]

Are you color-blind? [▽]

Mention any methods (such as mathematical theorems and diagrams you drew/used), material (such as pen and paper) or tools (such as calculator) you used to get your answer to the mammography, the cab and the economics course problems:

[ ]

Any other comments:

[ ]

Thank You!

Thank you very much for participating in our experiment. Now click on the submit button to complete this HIT.

[Submit]   (this will submit all pages)

Figure B.13: *Page 10 of 10 of Experiment 1 HIT for problem ordering Mam, Cab, Eco and visualization type V4.*

B.2.2 Qualitative Feedback After Experiment 1

As discussed in Section 6.5.6, three weeks after Experiment 1, all the participants were asked to complete a short follow-up questionnaire implemented as a one-page HIT on MTurk with some interactive features.

Figure B.14 illustrates how the HIT appears once it is loaded. If the subject clicked on "There was no diagram, just an icon and text" or "I don't remember at all", questions as in Figure B.15 were asked, otherwise questions as in Figure B.16 appeared. Once an answer to the questions was indicated on the Likert scales, the subject was asked to provide a reason for his or her choice as shown in the example in Figure B.17 and Figure B.18.

Figure B.14: *The qualitative feedback questionnaire HIT once it is loaded.*

diagram,
just an
icon and
text.

remember
at all.

Click here if you want to be reminded of the cab problem

Your answer (click an icon above to change):

→ **There was no diagram, just an icon and text.**

Please take a quick look at the diagrams above then select "There was no diagram" again.
Would you have **looked at** these diagrams if they had been provided to you?

Not at all		Somehow		Yes, very carefully
○1	○2	○3	○4	○5

Would you have used these diagrams **to get your answer** to the problems?

Not at all		Somehow		Would have used them a lot
○1	○2	○3	○4	○5

Additional comments or clarifications to the answers above:

Thank You!

Thank you very much for answering our follow-up questionnaire. Your feedback is invaluable to us.

Now click on the submit button to complete this HIT.

Submit (this will submit your HIT)

Figure B.15: *The bottom part of the qualitative feedback questionnaire HIT when the subject clicked on "There was no diagram, just an icon and text".* A similar display was provided when the subject clicked on "I don't remember at all".

diagram,
just an
icon and
text.

remember
at all.

Click here if you want to be reminded of the cab problem

Your answer (click an icon above to change):

cabs which the
witness identified
as Blue

Blue
cabs

Red
cabs

● a Blue cab   ● a Red cab
○ a cab which the witness identified as Blue

Overall, did you **look at** these diagrams?

Did not look at all		Somehow looked		Looked very carefully
○1	○2	○3	○4	○5

Overall, did you use these diagrams **to get your answer** to the problems?

Did not use them at all		Somehow used them		Used them a lot
○1	○2	○3	○4	○5

Additional comments or clarifications to the answers above:

Thank You!

Thank you very much for answering our follow-up questionnaire. Your feedback is invaluable to us.

Now click on the submit button to complete this HIT.

Submit   (this will submit your HIT)

Figure B.16: *The bottom part of the qualitative feedback questionnaire HIT when the subject clicked on visualization V4.* A similar display was provided when the subject clicked on any other visualization.

diagram,
just an
icon and
text.

remember
at all.

Click here if you want to be reminded of the cab problem

Your answer (click an icon above to change):

→ **There was no diagram, just an icon and text.**

Please take a quick look at the diagrams above then select "There was no diagram" again.
Would you have **looked at** these diagrams if they had been provided to you?

Not at all		Somehow		Yes, very carefully
○1	○2	○3	◉4	○5

Please explain **why** you would not have looked at them more carefully:

Would you have used these diagrams **to get your answer** to the problems?

Not at all		Somehow		Would have used them a lot
○1	○2	○3	◉4	○5

Please explain **how** you would have used them and **why** you would not have relied on them more:

Additional comments or clarifications to the answers above:

Thank You!

Thank you very much for answering our follow-up questionnaire. Your feedback is invaluable to us.

Now click on the submit button to complete this HIT.

Submit (this will submit your HIT)

Figure B.17: *An example of the questions asked just after the subject completes his Likert scale scores for the qualitative feedback questionnaire HIT when no visualization is selected for the first question.*

diagram,
just an
icon and
text.

remember
at all.

Click here if you want to be reminded of the cab problem

Your answer (click an icon above to change):

cabs which the
witness identified
as Blue

Blue
cabs

Red
cabs

● a Blue cab   ● a Red cab
○ a cab which the witness identified as Blue

Overall, did you **look at** these diagrams?

Did not look at all		Somehow looked		Looked very carefully
○1	●2	○3	○4	○5

Please explain **why** you essentially ignored them:

Overall, did you use these diagrams **to get your answer** to the problems?

Did not use them at all		Somehow used them		Used them a lot
○1	●2	○3	○4	○5

Please explain **how** you used them and **why** you did not rely on them more:

Additional comments or clarifications to the answers above:

Thank You!

Thank you very much for answering our follow-up questionnaire. Your feedback is invaluable to us.

Now click on the submit button to complete this HIT.

Submit   (this will submit your HIT)

Figure B.18: *An example of the questions asked just after the subject completes his Likert scale scores for the qualitative feedback questionnaire HIT when a visualization is selected for the first question.*

B.2.3 Experiment 2: Alternative Text Formats

As discussed in Section 6.6.3, the HITs were made up of four pages, including: a page with instructions; a page with the *Mam* problem presented using presentation type V0, classic text with no visualization, or V4, the classic text with visualization V4, or V4a, text with references to the diagram and visualization V4, or V4b, text with no numbers and visualization V4 (presentation types are discussed in Section 6.6.1; novel text format V4a and V4b are available in Table 6.6); a page with a catch question; a final page with questions asking the subjects whether they tried to compute an exact answer and how much they used the text and the diagram to solve the problem.

Each of the four presentation types were assigned to 120 different subjects. The following HIT illustrates presentation type V4b (text with no numbers and visualization V4).

Figure B.19: *Page 1 of 4 of Experiment 2 HIT for presentation type V4b.*

Page 2/4

Problem

Read the text below carefully and answer the question as best as you can.

You can use any method or tool to get your answer (pen and paper, calculator,...), or you can use no tool at all.

Your answer does not need to be exact.

A small minority of women at age forty who participate in routine screening have breast cancer.
A large proportion of women with breast cancer will get a positive mammography.
A small proportion of women without breast cancer will also get a positive mammography.

This is illustrated by the following diagram:

women with positive mammography

women with breast cancer

women without breast cancer

● a woman with breast cancer   ● a woman without breast cancer   ○ a woman with positive mammography
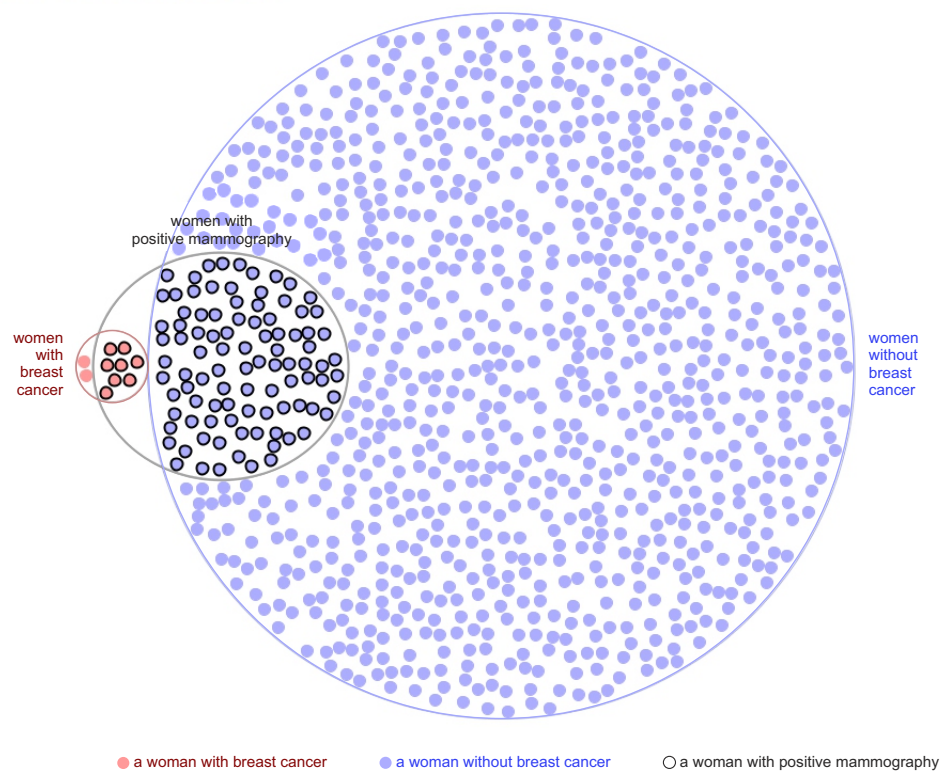
Please answer the following question:

Here is a new representative sample of women at age forty who got a positive mammography in routine screening.
How many of these women do you expect to actually have breast cancer?

Your answer: _____ out of _____

How confident are you in your answer?

No Confidence		Reasonable Confidence		Very High Confidence
○1	○2	○3	○4	○5

Next page (you cannot return to this page afterwards)

Figure B.20: *Page 2 of 4 of Experiment 2 HIT for presentation type V4b.*

Figure B.21: *Page 3 of 4 of Experiment 2 HIT for presentation type V4b.*

Page 4/4

Final Questions

Please take two more minutes to answer the questions below. Remember that you have to complete all the questions in the HIT to get your reward.

Q1. Did you try to compute an **exact answer** to the mammography problem? (as opposed to just estimating the answer)

[▼]

Recall the mammography problem was presented to you through a short piece of **text** followed by a visual **diagram**.

Q2. Did you use the information in the **text** to get your answer to the mammography problem?

Did not use it at all		Somehow used it		Used it a lot
○1	○2	○3	○4	○5

Q3. Did you use the **diagram** to get your answer to the mammography problem?

Did not use it at all		Somehow used it		Used it a lot
○1	○2	○3	○4	○5

[   ]

Q4. Did you previously start another HIT identical or similar to this but **without successfully submitting it** (e.g., because you experienced technical difficulties or because you gave up)? Your answer will **not** affect our decision to accept or reject this HIT.

[▼]

Any other comments:

[   ]

Thank You!

Thank you very much for participating in our experiment. Now click on the submit button to complete this HIT.

[Submit] (this will submit all pages)

Figure B.22: *Page 4 of 4 of Experiment 2 HIT for presentation type V4b.*

Appendix C

Further Reading

This appendix provides further reading on research areas related to this dissertation, including graph drawing, Bayesian reasoning, visualization and text, and Euler-like diagrams with glyphs.

C.1 Graph Drawing

Graphs are versatile and widely used in diverse application areas (e.g., social network analysis [Perer, 2010; Scott, 2012]; comparative genomics [Krzywinski et al., 2009]; brain connectivity analysis [Bullmore and Sporns, 2009]; stock market prices predictions [Shen et al., 2012]) to represent collections of data of various types and dimensions. The vertices can represent any required data elements, while the edges indicate relationships between the data elements [Bondy and Murty, 1976].

C.1.1 Aesthetics

Depicting the vertices as nodes and the edges as links could easily create a visualization (i.e., a node-link diagram) that is hard to understand and possibly unusable, due to visual clutter, various connections between vertices, edge crossings and other aesthetic features that could hinder comprehension [Bennett et al., 2007; Díaz et al., 2002]. Studies have been conducted to identify aesthetic features that should be avoided and others that can facilitate understanding. This led to the formulation and prioritization of a number of structural (e.g., minimizing the length of the paths, edge bends and edge crossings; maximizing orthogonality) and semantic (e.g., using equally-sized vertices) rules and conventions for drawing such graphs [Bennett et al., 2007] (also [Sugiyama, 2002]).

Most of these aesthetic criteria are consistent with perceptual principles [Bennett et al., 2007], such as the Gestalt principles [Koffka, 1935] and Norman's emotional design framework [Norman, 2004] (also [Norman, 1988, 1991, 2002]). Studies have also been conducted to evaluate their effectiveness in helping potential users carry out specific tasks on such graph drawings (e.g., [Dwyer et al., 2009; Huang, 2013; Purchase et al., 2001, 2002, 2007, 2012; Purchase, 2000]). Formal metrics to quantify the aesthetic quality of a graph drawing [Purchase et al., 2002] and cognitive measures to quantify the cognitive effort required to use a graph drawing with specific aesthetic features [Ware et al., 2002] (also

[Huang et al., 2006, 2008, 2009]) are also available. Such metrics could be used to define the cost function of automatic graph drawing algorithms, such as those employing simulated annealing, genetic algorithms or evolutionary algorithms.

C.1.2 Layout Methods

Various automatic techniques have been developed to draw graphs [Di Battista et al., 1994, 1999b; Díaz et al., 2002; Gibson et al., 2012; Kaufmann and Wagner, 2001; Sugiyama, 2002]. The majority strive to satisfy most of the defined aesthetic criteria to ensure that their generated layouts are comprehensible and usable (e.g., [Coleman and Parker, 1996; Tamassia et al., 1988]), but often, this is not possible, as various aesthetic criteria could be in conflict, and complex and computationally expensive algorithms would be required to handle multiple aesthetic features simultaneously [Di Battista et al., 1999b; Díaz et al., 2002]. Also, the precedence of aesthetic features varies between application areas [Di Battista et al., 1999b; Díaz et al., 2002]. So, techniques that produce different layouts based on specific aesthetic criteria with different precedence are available [Di Battista et al., 1994] (see Table 1 of Gibson et al.'s [2012] survey for a comparison of different algorithms).

Examples of graph layouts generated by currently available algorithms, include: straight-line layouts (e.g., [Eades et al., 1997]), so each edge is depicted as a one-line segment; planar layouts (e.g., [Nishizeki and Rahman, 2004]), so none of the edges cross; orthogonal layouts (e.g., [Biedl and Kant, 1998]), so each edge is depicted as one or a set of connected horizontal or vertical line segments; circular layouts (e.g., [Krzywinski et al., 2009]), in that the depicted vertices are placed on a circle; hierarchical layouts (e.g., [Eades et al., 1996]), in that the depicted vertices are positioned in hierarchical arranged levels (where levels are depicted as parallel lines) to visualize the flow of a directed graph.

Graph drawing algorithms (e.g., using simulated annealing [Coleman and Parker, 1996] or a genetic algorithm [Hobbs and Rodgers, 1998] for multi-criteria optimization; using evolutionary algorithms [Utech et al., 1998]) can be complex and computationally expensive. So, lately, force-directed algorithms (previously used for circuit layouts [Fisk et al., 1967]) have been used [Gibson et al., 2012] to generate comprehensible layouts in relatively fast time [Kobourov, 2012].

C.2 Bayesian Reasoning

C.2.1 Applying Bayes' Theorem Using Probabilities or Frequencies

Given two events A and B, the probability that A occurs given that B already happened is denoted as $P(A|B)$ and can be computed using Bayes' [1763] theorem as follows

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

where $P(A)$ is the probability that A occurs, $P(B)$ is the probability that B occurs, and $P(B|A)$ is the probability that B occurs given that A already happened.

Let us now assume that *A* and *B* are respectively *H* (a hypothesis) and *D* (some data obtained from an observation). When we have such events, we are often interested to know the likelihood that the hypothesis is true given the data. So we need to find the conditional probability *P(H|D)*, which as we have seen above can be computed by applying Bayes' theorem

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D)}$$

$$= \frac{P(D \mid H)P(H)}{P(D \mid H)P(H) + P(D \mid \overline{H})P(\overline{H})}$$

(C.2)

where $\overline{H}$ is the complement of *H*, *P(H)* is the probability that *H* is true, *P(D)* is the probability that *D* is obtained, and *P(D|H)* is the probability that *D* is obtained given that *H* is true [Gigerenzer and Hoffrage, 1995].

Similarly, Bayes' theorem can be used to solve the classic mammography Bayesian problem:

The probability that a woman at age 40 has breast cancer is 1%. According to the literature, the probability that the disease is detected by a mammography is 80%. The probability that the test misdetects the disease although the patient does not have it is 9.6%.

If a woman at age 40 is tested as positive, what is the probability that she indeed has breast cancer?

[Eddy, 1982]

If in Equation C.2, *H* and $\overline{H}$ are respectively *breast cancer* and *no breast cancer* and *D* is *positive mammography* [Gigerenzer and Hoffrage, 1995], then

$$P(breast\ cancer \mid positive\ mammography) = \frac{(0.8)(0.01)}{(0.8)(0.01) + (0.096)(0.99)}$$

$$= 0.078$$

$$= 7.8\%$$

So, the probability that a woman at age 40 indeed has breast cancer given that she is tested as positive by a mammography is 7.8%.

The probabilities in the above problem could be replaced with natural frequencies:

10 out of every 1,000 women at age forty who participate in routine screening have breast cancer. 8 of every 10 women with breast cancer will get a positive mammography. 95 out of every 990 women without breast cancer will also get a positive mammography.

Here is a new representative sample of women at age forty who got a positive mammography in routine screening. How many of these women do you expect to actually have breast cancer?

[Gigerenzer and Hoffrage, 1995]

From the above, the following is known: the number of women who have got a positive mammography and indeed have breast cancer (i.e., 8); the number of women who have also got a positive mammography but do not have breast cancer (i.e., 95). Knowing these two quantities, all the other quantitative values in the problem can be ignored and the answer can be computed as follows

$$P(breast\ cancer \mid positive\ mammography) = \frac{8}{8 + 95}$$
$$= 0.078$$
$$= 7.8\%$$

This answer is the same as that obtained when probabilities are used [Gigerenzer and Hoffrage, 1995].

Hence, when natural frequencies are used, the conditional probability $P(H \mid D)$ can be computed as

$$P(H \mid D) = \frac{\left|D \cap H\right|}{\left|D \cap H\right| + \left|D \cap \overline{H}\right|} \tag{C.3}$$

where $\left|D \cap H\right|$ is the number of cases consistent with the data and the hypothesis, while $\left|D \cap \overline{H}\right|$ is the number of cases consistent with the data but not the hypothesis [Gigerenzer and Hoffrage, 1995].

Both Equation C.2 and Equation C.3 are mathematically equivalent to Bayes' [1763] theorem [Gigerenzer and Hoffrage, 1995]. However, the above computations using both equations demonstrate Gigerenzer and Hoffrage's [1995, pp. 685-87] claim that it is "simpler" to solve a Bayesian problem that uses natural frequencies than one that uses probabilities. This is so, as the former requires fewer calculations, less complex computation, fewer pieces of information (only the hit rate and the false alarm rate), and less thought on how the base rate should be used as it is inherently carried on to the hit rate and false alarm rate.

C.2.2 Probabilities versus Natural Frequencies

Bayesian problems that present their quantitative data using natural frequencies instead of probabilities reduce the fallacy [Gigerenzer and Hoffrage, 1995, 1999; Hoffrage et al., 2002], with a typical improvement that goes from 16% to 46% of correct answers [Gigerenzer and Hoffrage, 1995].

A possible reason for this improvement is that using natural frequencies, information about the base rate (in this case, 1% or "10 of every 1,000") is inherently carried on to both the hit rate (in this case, 80% or "8 of every 10") and the false alarm rate (in this case, 9.6% or "95 out of every 990"), both of which are important in solving the Bayesian problem and thus, information about the base rate is not lost [Hoffrage et al., 2000]. This would also reduce the amount of required computation to obtain an answer [Gigerenzer and Hoffrage, 1995; Hoffrage et al., 2000] and the complexity of the computations which are carried out on natural frequencies, that is natural countable numbers, instead of probabilities, that is real continuous numbers and fractions [Gigerenzer and Hoffrage, 1995] (see Appendix C.2.1 to compare the required computation to solve a problem when using probabilities and when using natural frequencies). Moreover, natural frequencies explicitly indicate the reference class [Gigerenzer et al., 2007], so from the first line of the mammography problem ("10 out of every 1,000 women …"), we immediately know that "10" is referring to and is part of the class of "1,000 women". It is also thought

that humans have a specific natural affinity to natural frequencies and that humans have evolved ways how to handle information in this format which has long been used before the rise of probability theory [Cosmides and Tooby, 1996; Gigerenzer and Hoffrage, 1995, 1999; Gigerenzer, 1998].

However, some argue that natural frequencies can be effective only because they provide a possible representation that facilitates the construction of a mental map that clarifies the nested-set relations of the problem, that is, how the base rate, the hit rate and the false alarm rate are related to each other [Mellers and McGraw, 1999; Moro, 2010; Sloman et al., 2003], which benefit had been conjectured much earlier [Tversky and Kahneman, 1983]. Others rejected the view that humans have cognitive abilities to handle natural frequencies and not probabilities [Girotto and Gonzalez, 2002] and disputed the possible benefits of natural frequencies over probabilities, emphasizing instead the importance for representations to clarify the nested-set relations of the problem [Evans et al., 2000; Girotto and Gonzalez, 2001, 2002]. Another study among physicians suggests that carrying out Bayesian inferences and computing the outcome of an event that occurred in the past using natural frequencies is more effective than using probabilities and yields better performance than using past experiences to predict the outcome of the event in the future in natural frequencies [Obrecht et al., 2012].

Despite these debates, it is clear that representations of Bayesian problems should clarify their structure and the relationships of the different numerical quantities. If using natural frequencies can help to accomplish this objective, irrespective of the reason why this is true, then their use in representations is rational. However, the success rate in solving Bayesian problems when information is provided textually using natural frequencies could still be as low as 46% [Gigerenzer and Hoffrage, 1995] and so, other representations and aids should be studied for further improvement and facilitation.

C.2.3 Facilitating Bayesian Reasoning

A few techniques to solve Bayesian problems, besides applying the Bayes' theorem [Bayes, 1763], include: the use of heuristics [Bar-Hillel, 1983; Gigerenzer et al., 1999; Gigerenzer and Selten, 2002; Kahneman et al., 1982; Tversky and Kahneman, 1974], the construction of mental models [Johnson-Laird et al., 1999], Bayesian [Gigerenzer and Hoffrage, 1995, p. 689] and non-Bayesian [Gigerenzer and Hoffrage, 1995, p. 694] cognitive algorithms, and inferencing mechanisms over causal models [Krynski and Tenenbaum, 2007; Tversky and Kahneman, 1980].

Diagrams are known to facilitate reasoning [Bauer and Johnson-Laird, 1993; Sedlmeier, 1999] and have been used extensively for the communication of numeric quantitative data [Macdonald-Ross, 1977] and for visualizing uncertainty (e.g., [Pang et al., 1997; Zuk, 2008]) to a widely diverse audience [Spiegelhalter et al., 2011], particularly for risk communication and comprehension [Ancker et al., 2006; Garcia-Retamero and Galesic, 2013; Lipkus, 2007; Schirillo and Stone, 2005; Visschers et al., 2009] to, for instance, communicate earthquake [Bostrom et al., 2008] and climate change risks [Sterman, 2011], promote healthier lifestyles (e.g., [Cox et al., 2010; Garcia-Retamero and Cokely, 2011]), make informed decisions about different possible therapies (e.g., [Elting et al., 1999; Goodyear-Smith et al., 2008; Waters et al., 2007; Zikmund-Fisher et al., 2008a, b]) based on statistical rather than anecdotal information (e.g., [Fagerlin et al., 2005]), communicate information to non-natives who have difficulties with the language (e.g., [Garcia-Retamero and Dhami, 2011]). Bayesian reasoning can also be facilitated with visual representations, so information is sometimes provided textually and visually.

In a study that tested textual representations using probabilities and natural frequencies as the mammography example we previously discussed, a participant voluntarily drew a "pictorial analog" (Figure C.1) of the Bayes problem to help him carry out the required inferences and solve the problem [Gigerenzer and Hoffrage, 1995]. Actually, in his original essay just after providing the definitions and proofs of propositions, Bayes described the problem he was attempting to solve with his theorem using an analogue of a table on which balls are thrown and illustrated this with a diagram [Bayes, 1763], which analogue later became known as Bayes' billiard table. A study confirms that when Bayes' theorem is introduced to students through visualizations, students learn faster and report higher temporal stability than without a visualization [Sedlmeier and Gigerenzer, 2001]. However, prior training is not always possible, particularly in cases such as medical patients, juries, and news readers who are given access to growing amounts of open statistical and scientific data [Hall, 2011]. A few studies were conducted to assess the immediate benefits of visualizations (e.g., [Brase, 2009; Cole and Davidson, 1989; Cole, 1989; Sloman et al., 2003]) and though they confirm that visualizations facilitate reasoning, it is still unclear which is the most effective representation for Bayesian reasoning. Moreover, studies have been carried out on small populations with a specific background (usually highly-focused university students; e.g., [Brase, 2009; Gigerenzer and Hoffrage, 1995; Sloman et al., 2003]), making it difficult to generalize their findings to a more diverse population of laypeople of various backgrounds and ages. These problems also pertain to the more general area of risk communication where no solid recommendations on how visual representations can effectively and transparently communicate statistics to inform rather than manipulate or persuade [Gigerenzer et al., 2007; Nelson et al., 2009] have not been formulated [Visschers et al., 2009] except for Spiegelhalter et al.'s "broad conclusions" [2011, p. 1399 Box 1] .
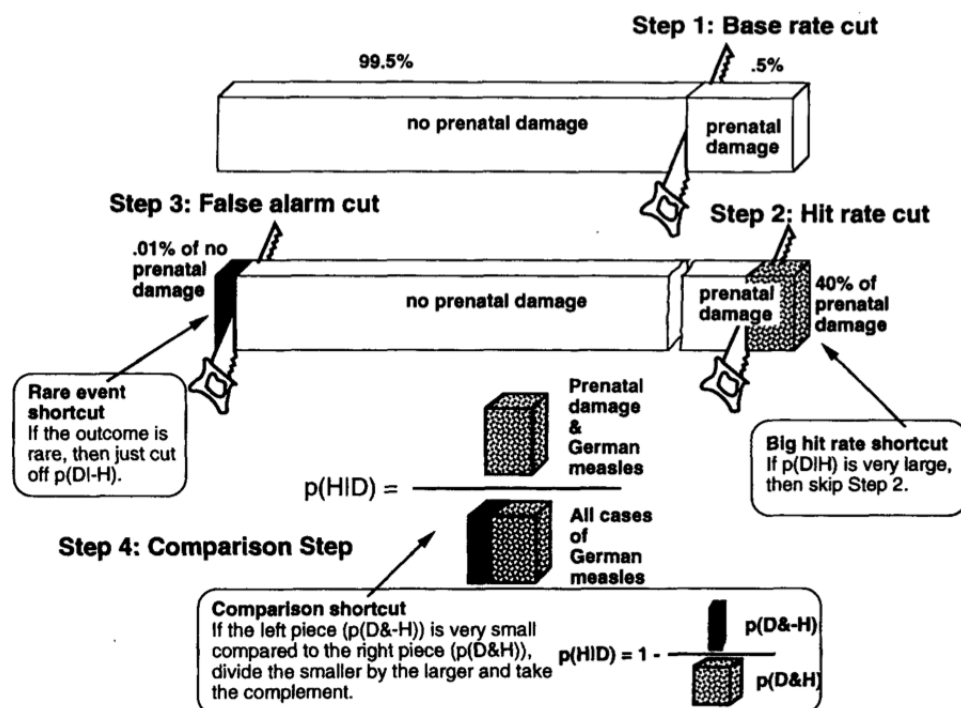

Figure C.1: *A pictorial analogue known as "beam cut" of a Bayesian problem drawn voluntarily by a participant during a study to facilitate inferencing.* This was drawn for the German measles problem. *Source*: [Gigerenzer and Hoffrage, 1995]–Figure 2

The visual representations that have been considered for Bayesian reasoning include: contingency tables [Cole and Davidson, 1989; Cole, 1989] (Figure C.2A), detection bars [Cole and Davidson, 1989; Cole, 1989](Figure C.2B), signal detection curves [Cole and Davidson, 1989; Cole, 1989] (Figure C.2C), "beam cut" pictorial analogues [Gigerenzer and Hoffrage, 1995] (Figure C.1), Bayesian Boxes [Burns, 2003, 2004a, b] (Figure C.2D), probability space diagrams [Cheng and Pitt, 2003] (Figure C.2E), frequency trees [Sedlmeier and Gigerenzer, 2001] (Figure C.2F), probability trees [Sedlmeier and Gigerenzer, 2001] (Figure C.2G), frequency grids (e.g., [Brase, 2009]) and Euler diagrams (e.g., [Sloman et al., 2003]). Trees [Yamagishi, 2003] and roulette-wheel diagrams [Ichikawa, 1989; Tubau, 2008; Yamagishi, 2003] (Figure C.3) have also been used to facilitate Bayesian reasoning for a different fallacy (other than the base rate fallacy), that of the 'Problem of Three Prisoners' [Falk, 1992; Lindley, 1994; Mosteller, 1987; Shimojo and Ichikawa, 1989] and the mathematically equivalent Monty Hall problem [Selvin, 1975]. However, though all of these representations attempt to visualize the computational structure of the problem, most are difficult to understand, thus requiring training prior to use.
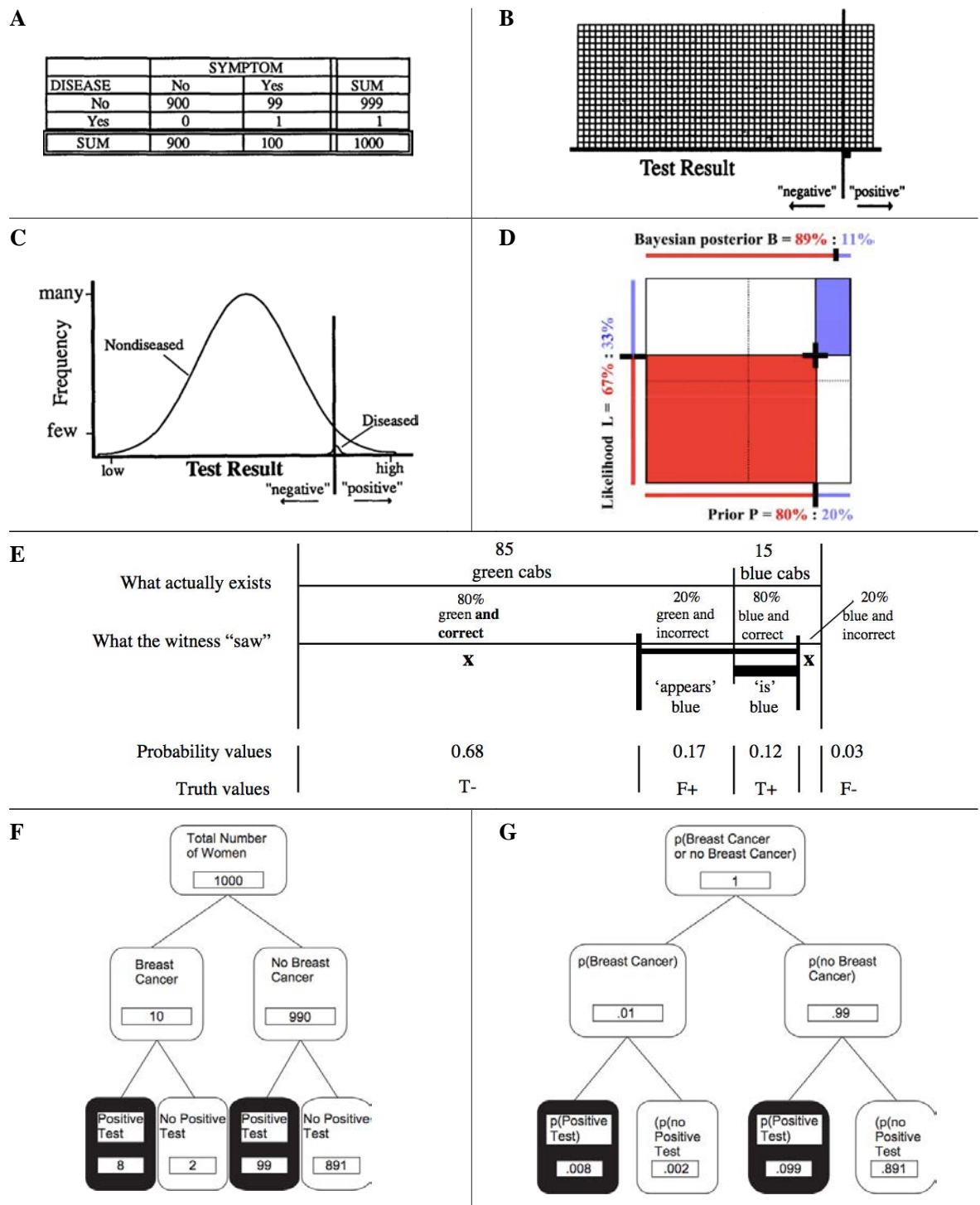
Figure C.2: *Different visual representations that have been considered for Bayesian reasoning to avoid the base rate fallacy (excluding Euler diagrams and frequency grids).* (A) A contingency table (for the disease diagnosis problem [Casscells et al., 1978]). *Source*: [Cole, 1989]—Table 1 (B) A detection bar (for the disease diagnosis problem [Casscells et al., 1978]). *Source*: [Cole, 1989]—Figure 2 (C) A signal detection curve (for the disease diagnosis problem [Casscells et al., 1978]). *Source*: [Cole, 1989]—Figure 1 (D) Bayesian Boxes (for the King and Queen card problem [Burns, 2004b]). *Source*: [Burns, 2004b]—Figure 1 (E) A Probability Space diagram (for the cab problem [Tversky and Kahneman, 1982]). *Source*: [Cheng and Pitt, 2003]—Figure 6 (F) A frequency tree (for the mammography problem [Gigerenzer and Hoffrage, 1995]). *Source*: [Sedlmeier and Gigerenzer, 2001]—Figure 7 (G) A probability tree (for the mammography problem [Eddy, 1982]). *Source*: [Sedlmeier and Gigerenzer, 2001]—Figure 11
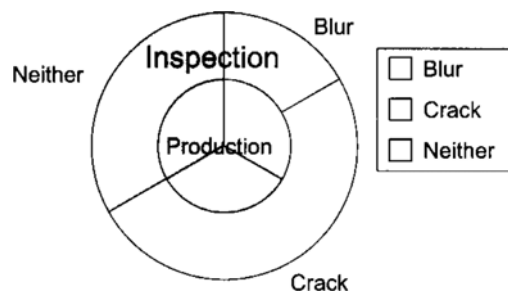
Figure C.3: *A roulette-wheel diagram for Bayesian reasoning to facilitate the 'Problem of Three Prisoners' and the Monty Hall problem.* Representing an isomorph of the 'Problem of Three Prisoners' [Lindley, 1994; Mosteller, 1987; Shimojo and Ichikawa, 1989] named the Gemstone Problem [Yamagishi, 2003]. *Source*: [Yamagishi, 2003]—Figure 2

C.2.4 Iconic Visualizations

Kellen et al. [2007] hypothesized that an iconic visual representation, that is a visualization with discrete objects like Figure C.4, can aid users with different spatial abilities solve Bayesian problems. However, they did not evaluate their proposed visual representation.
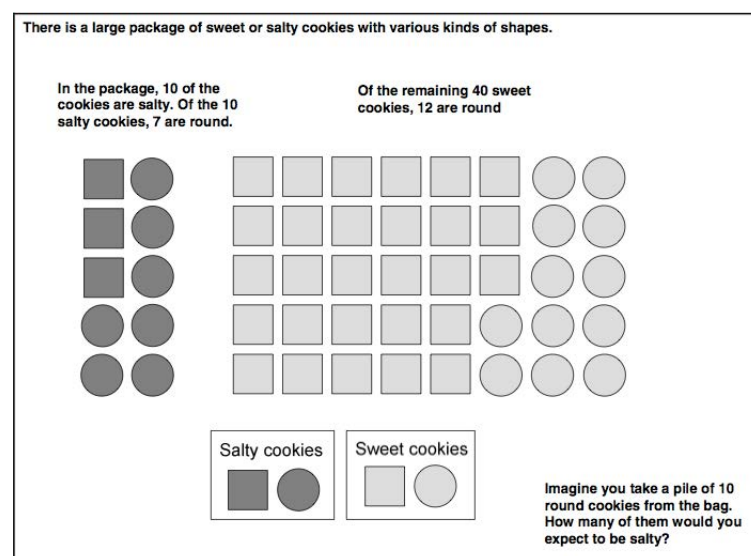

Figure C.4: *Kellen et al.'s [2007] proposed iconic visualization. Source*: [Kellen et al., 2007]–Figure 2

C.2.5 Frequency Grids with Tree Diagrams

Spiegelhalter et al. [2011] argued that having a frequency grid with a tree diagram, as in Figure C.5 (depicting the mammography problem we discussed earlier, but using different quantitative values), could be beneficial, as it clearly illustrates how the quantitative values are relate, how portions of the sampled population representing sub-classes relate to the entire sampled population and how the Bayesian inferencing is carried out. However, no studies have been conducted. Both diagrams (the frequency grid and the tree diagram) illustrate the same information and so, showing duplicate and redundant data might be more confusing and would required further time to be processed [Tufte, 1990]. Spiegelhalter et al. [2011] also comment that interactions visualization can assist the user in their reasoning and thus, they designed an interactive and animated version of the visualization in Figure C.5

([Spiegelhalter et al., 2011]—Movie S3). They also designed an interactive version of Nightingale's iconic 'rose' or 'coxcomb' 'Diagram of the Causes of Mortality in the Army in the East' [Nightingale, 1858] (Figure 2.3) and added to it an additional interactive visualization representing the same data as frequency grids ([Spiegelhalter et al., 2011]—Movie S1).
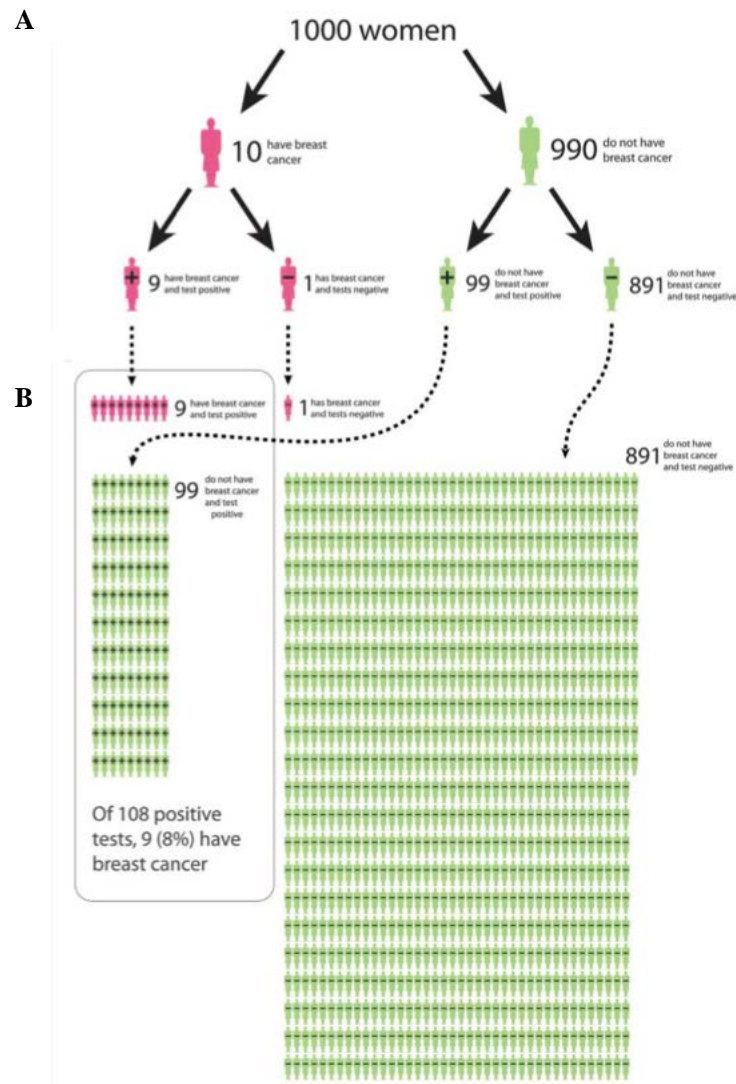

Figure C.5: *A frequency grid with a tree diagram for the classic mammography Bayesian problem.* (A) A tree diagram with (B) a frequency grid, illustrating the data published by the UK Breast Cancer Screen Programme in 2011 [Cancer Research UK, 2012]. This depicts the classic mammography problem (probability format [Eddy, 1982]; natural frequencies format [Gigerenzer and Hoffrage, 1995]) but using different quantitative data, that is 1% or "10 out of 1000" base rate, 90% or "9 out of 10" hit rate, 10% or "99 out of 990" false alarm rate, so the answer is 8% or "9 out of 108". *Source*: [Spiegelhalter et al., 2011]–Figure 4; an interactive and animated version is also available [Spiegelhalter et al., 2011]–Movie S3

C.2.6 Frequency Grids with Interactive Features

Tsai's [2012] studied frequency grids as Figure C.6 with interactive features. The main rational is that, in contrast to passive visualizations which are only provided with the textual information, these interactive visualizations allow the readers to engage in the construction process of the visualization. Using the checkboxes at the side, the user can include (or exclude) the different pieces of information to

(or from) the diagram as they read it off the text and while they are actively reasoning about the problem. Initially, all the boxes in the grid are grey and all the checkboxes except for the first one are unchecked (so the grid only illustrates the entire sampled population with no sub-groups). The author also predicted that hiding the yellow and the green boxes, so only the yellow and the dark green striped boxes are visible (the rest would be coloured in grey), would make the answer to the Bayesian problem more transparent and the area of these sections easily comparable.
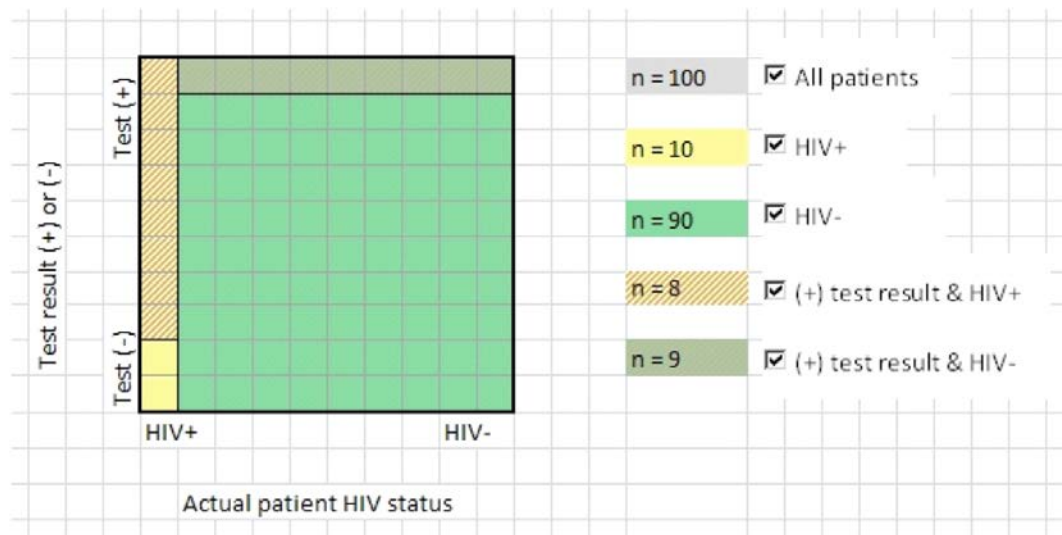

Figure C.6: *A frequency grid with interactive features for the HIV diagnosis problem.* This was designed and tested by Tsai [2012] and works in Microsoft Excel 2007 using VBA macros. This example depicts the disease diagnosis problem. Out of a population of 100, 10 are HIV positive (yellow boxes—base rate) and out of which, 8 get a positive HIV test (yellow with stripes boxes—hit rate). Out of the 90 who are not HIV positive, 9 also get a positive HIV test (dark green with stripes boxes—false alarm rate). So the likelihood that a patient is HIV positive, given a positive HIV test is 8 out of 17 or 47%. The checkboxes at the side can be used to include or exclude different groups from the diagram. So, for instance, the dark green with stripes boxes will change to light green (no stripes) when the last checkbox labelled as '(+) test result & HIV' is unchecked. *Source*: [Tsai, 2012, p. 31]–Figure 6

Over six Bayesian problems (adaptations from Gigerenzer and Hoffrage's [1995] problems), Tsai reported a success rate of 73% of 15 participants for the interactive frequency grid and a success rate of 49% of 15 participants for text alone using natural frequencies (participants were from various Cambridge, Massachusetts, USA universities and local communities with age range of 18 to 32 years and were paid $10 for the one-hour session; though they completed college education, none had an understanding of Bayes' theorem and conditional probabilities).

This result for the interaction frequency grid is similar to those of Cosmides and Tooby [1996] and Sedlmeier and Gigerenzer [2001] for passive frequency grids, but are lower than those obtained by Cosmides and Tooby [1996] for their actively drawn frequency grids. Thus, such interactive visualizations might not be the best option to actively engage the reader and yet they might not be practical in the real-world, as participants had to be instructed ("a short explanation lasting 1-2 minutes" [Tsai, 2012, p. 34]) prior to the experiment. Tsai also reported an improved performance using similar interactive frequency grids compared to only text for more complex problems known as 'chains of reasoning' that involve multiple sequential Bayesian inferences and the integration of various independent pieces of information in the understanding of the likelihood that an event will occur [Schum, 1991, 1999].

C.2.7 The Main Concept Behind Frequency Grids

The idea of frequency grids have long been used for various applications ([Macdonald-Ross, 1977, pp. 381-91]—brief review). In around 1930s, Neurath created the international picture language Isotype [Neurath, 1936] (see also [Neurath and Kinross, 2009; Neurath and Neurath, 2012; Neurath, 2010]) as an education tool to represent facts visually using objects and simple figures, with the primary rule stating that quantities should be represented as a collection of same sized pictograms rather than enlarged ones. This is so, as Neurath believed that, in contrast to repeated pictograms which can be counted, size cannot be compared and judged accurately and could be misleading if it is unclear how size is encoded (e.g., as length or area) [Neurath, 1936]. Figure C.7 is an example that used Isotype to illustrate women's employment worldwide in the book, 'Women and Work' [Williams, 1945].
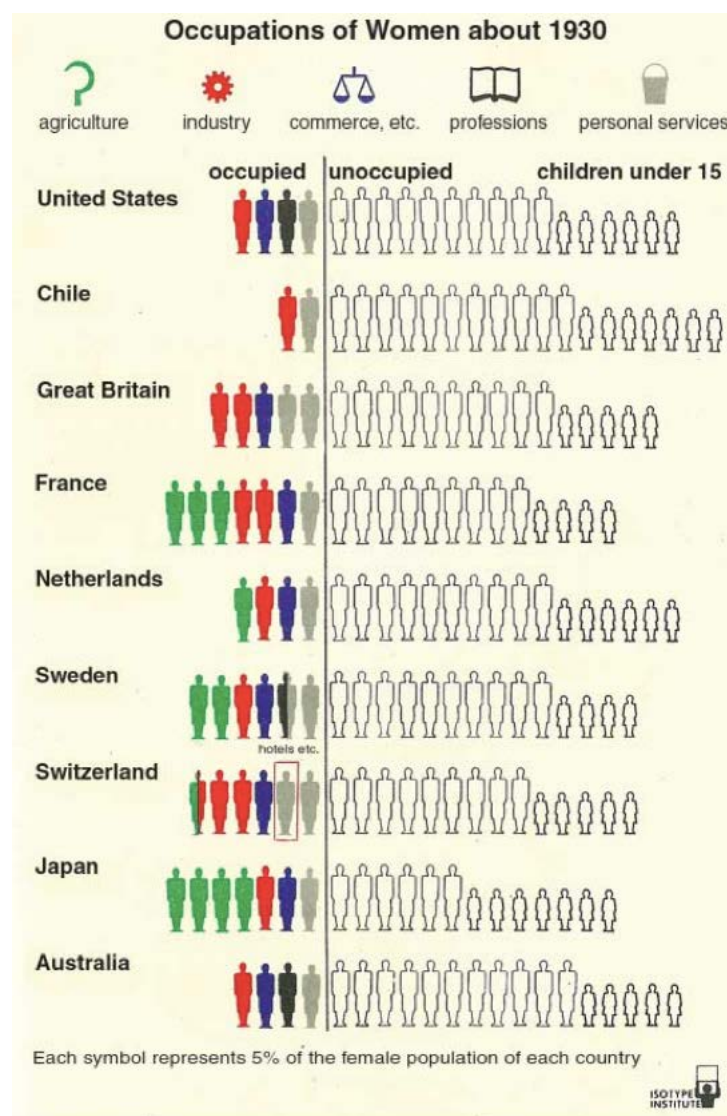

Figure C.7: *A visualization similar to a frequency grid drawn using the picture language Isotype to illustrate women's employment worldwide in 1930.* This was originally published in the book 'Women and Work' [Williams, 1945] and drawn using Isotype [Neurath, 1936] as most of the other diagrams in the book. *Source*: [Neurath and Neurath, 2012]

C.2.8 Frequency Grid in the General Area of Risk Communication

Frequency grids have also been used to facilitate reasoning in the more general area of risk communication and comprehension (e.g., [Dolan and Iadarola, 2008; Elmore and Gigerenzer, 2005; Elting et al., 1999; Kurz-Milcke et al., 2008; Price et al., 2007; Schapira et al., 2001; Waters et al., 2007; Zikmund-Fisher et al., 2008a, b]) even for elderly [Fuller et al., 2001, 2002]. These visualizations are considered more "attractive" than representations with only text and numbers [Gaissmaier et al., 2012], are rated "favorably" [Hawley et al., 2008], and are regarded as the most effective representation in communicating risk [Elting et al., 1999; Hawley et al., 2008; Schapira et al., 2001; Zikmund-Fisher et al., 2008a] be it when numerical information should be read from the representation (referred to as verbatim knowledge) or when the essence of the presented information should be understood (referred to as gist knowledge) [Hawley et al., 2008] (when compared to other visualizations such as pie charts and bar graphs and respective adaptations such as "sparkplug" and "clock" [Hawley et al., 2008]) and a way of reducing related biases such as the denominator neglect [Garcia-Retamero et al., 2010]. They also reduce the effect of anecdotal information in making informed decisions about a medical therapy [Fagerlin et al., 2005]. Possible successful features of frequency grids is that they clearly illustrate part-to-whole relationships (e.g., the number of the women who have breast cancer in the sampled populations compared to all the women in the sampled population) [Ancker et al., 2006; Spiegelhalter et al., 2011] and allow users to reason about the problem as if they are a character in the story (e.g., a woman carrying out the mammography test) [Kurz-Milcke et al., 2008].

In contrast to the more specific area of Bayesian reasoning, the design of frequency grids for risk communication in general has been studied thoroughly ([Ancker et al., 2006, pp. 610-12]—brief review). It was shown that visualizations made up of two simple grids (each containing quantities about two same concepts but differ by a third factor) are perceived faster than one complex grid (illustrating the same information as in the two simple grids) [Price et al., 2007] and those containing the frequency grid (with all the quantitative information) and an adjacent magnified view of the section of the grid with the smaller quantities facilitate reasoning about events that are less likely to occur [Dolan and Iadarola, 2008]. Grids with 10 or 100 glyphs are often preferred and considered easier to handle than ones with more glyphs (even though portions of a glyph could be highlighted to represent the required quantity) [Schapira et al., 2001; Viscusi, 1992], but still, confidence in the presented information might be lost when grids with as few as 10 glyphs are used [Schapira et al., 2001] due to human intuition and the 'Law of Large Numbers' that larger samples provide a mean that is closer to the expected value than smaller ones and are thus considered more reliable [Bernoulli, 1713; Sedlmeier and Gigerenzer, 1997]. A study suggests that grids with more glyphs (e.g., 1000 glyphs) could perceive risk as being smaller as unhighlighted glyphs get more attention than the highlighted ones [Schapira et al., 2001] (though this might not be fully consistent with the ratio-bias phenomenon whereby a probability of, for instance, 1 in 20 could be perceived smaller than 50 in 1000 [Denes-Raj et al., 1995]).

Some also warn about the effect of inappropriate designs. For instance, frequency grids with random layouts are ineffective, make comparisons and judgement of the occurrence of events difficult, and could mislead viewers in overestimating the size of large proportions [Ancker et al., 2011]. Also, frequency grids that show just part of the sampled population, so that, in a case where a therapy should be chosen, the grid contains glyphs for just the patients who were cured instead of having highlighted and unhighlighted glyphs to respectively represent patients who were and were not cured, could be misleading as the attention is drawn to patients who were cured, while information about those who were not cured is ignored [Stone et al., 2003].

C.2.9 Frequency Grids For Different Abilities

A study in risk communication showed that the level of visual similarity of the actual glyphs to the objects they are representing (e.g., using stick figures, as in Figure 2.10, or a photograph of a person instead of a simple geometric shape like a circle or square, as in Figure 2.9, to show that each glyph is representing a person) does not have an effect on comprehension and recall, but the graph literacy of the user (i.e., the user's ability to understand and handle graphical information [Aldrich and Sheppard, 2000]) does [Gaissmaier et al., 2012]. So, those with high graph literacy had higher comprehension and recall scores when a graphic composed of a number of glyphs was provided and those with low graph literacy had better scores when numbers with some text was given [Gaissmaier et al., 2012]. Besides graph literacy, numeracy abilities (i.e., the user's ability to understand and handle numerical information [Peters, 2012], which abilities are related to various demographic factors such as age, sex and country [Galesic and Garcia-Retamero, 2010; Gigerenzer, 2010], but not necessarily the level of education [Brown et al., 2011; Estrada et al., 1999; Lipkus et al., 2001; Sheridan et al., 2002]) can also affect comprehension and reasoning [Fagerlin et al., 2007a; Nelson et al., 2008; Peters et al., 2006, 2007; Reyna et al., 2009]. Studies show that while those with high numeracy abilities benefit from any visualization in contrast to only numbers with text, those with low numeracy benefit mostly from visualizations as frequency grids [Hawley et al., 2008]. Thus, though using such visualizations the success rate of those with low numeracy is less than that of those with high numeracy [Galesic et al., 2009; Hawley et al., 2008], frequency grids are regarded to be the best representation for risk communication for users with different numeracy abilities [Hawley et al., 2008].

Different abilities and skills are required for a visual representation to facilitate comprehension and reasoning [Galesic and Garcia-Retamero, 2011; Spiegelhalter et al., 2011]. However, not everyone has the same abilities (e.g., [Galesic and Garcia-Retamero, 2010; Gigerenzer, 2010]) and thus needs (e.g., those with high numeracy are more likely to use the base rate [Obrecht and Chesney, 2013] and gain more insight from numbers [Peters et al., 2006]). So, either different visual representations are used for different levels of abilities [Spiegelhalter et al., 2011, p. 1399 Box 1] or visualizations that meet the abilities of different readers statically or interactively are designed. Interactive visualizations could assist readers and adapt to their needs and preferences [Strecher et al., 1999], but might not be accessible to all due to software incompatibility and the users' abilities in using the technology [Spiegelhalter et al., 2011], and users might not be willing to spend extra time and effort in understanding how the technology works, since often (e.g., [Tsai, 2012]) users are instructed prior to use. Also, adapting the visualization could help, but is not practical in the real-world [Hawley et al., 2008]. So static visualizations that meet the needs of different users might be preferred.

C.3 Visualization and Text

Visualizations are often added to text to aid readers understand the presented textual information. However, this is possible only when the visualization and the text are linked appropriately. For instance, Minard's [1869] flow map (Figure C.8) is a classic in information visualization that "exemplifies many of the fundamental principles of analytical design" [Tufte, 2006, p. 126] and illustrates how to "Completely integrate words, numbers, images, diagrams" [Tufte, 2006, p. 126] (further discussed in Section 2.6).
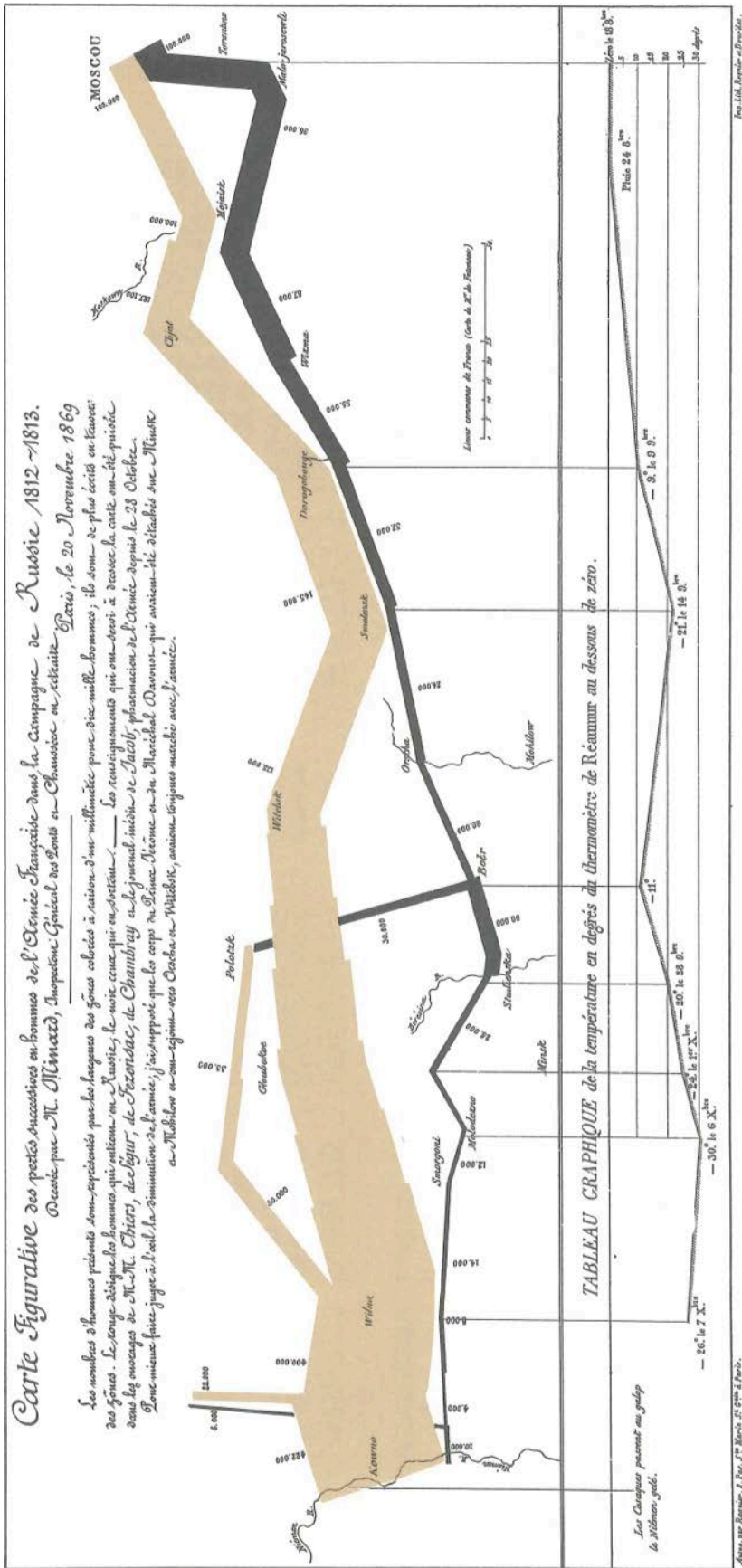
Figure C.8: *Minard's 1869 flow map illustrating the loss of men in Napoleon's army during their invasion of Russia between 1812 and 1813*. Item 28 in Minard's [1869] portfolio of statistical maps. Tufte describes this as "the best statistical graphics ever" [2006, p. 122]. Source: [Tufte, 2006, p. 136]

C.3.1 In the Past

Looking back in history, we can see different techniques that have been used. Leonardo da Vinci carefully placed drawings within blocks of relevant text such that the two are seen holistically (e.g., Figure C.9A [da Vinci, 1956, p. 234 para. 827]), while Galileo Galilei used drawings as sentence elements instead of words (e.g., Figure C.9B [Galilei, 1613, p. 25]). René Descartes, in the first edition of his book 'Principia Philosophiae' [Descartes, 1644], repeated a specific diagram 11 times to prevent the reader from flipping over pages to relate the text and the diagram. Isaac Newton, in his book 'Principia' [Newton, 1687], neatly integrated the various physics diagrams within the text (e.g., Figure C.9C) and often used labels to refer to instances in the diagrams and replicated diagrams (such as the one on the right in Figure C.9C) multiple times in the book to ensure that the diagram is always closely integrated to the text that refers to it. This was not the case with of the first edition of Newton's fundamental physics book 'Opticks' [Newton, 1704] (containing major discoveries such as the colours of light and the designs of the refractor telescope). Though a third of the text directly refers to the 55 diagrams in the book, all the diagrams were printed on separate flaps of paper that were later bound to the rest of the book, so that reader had to first understanding the label coding of the flaps (e.g., "Book I. Part II. Plate IV. Fig. 16") and then flip back and forth the pages of the book to link the text to the diagram [Tufte, 2006, pp. 110-11].
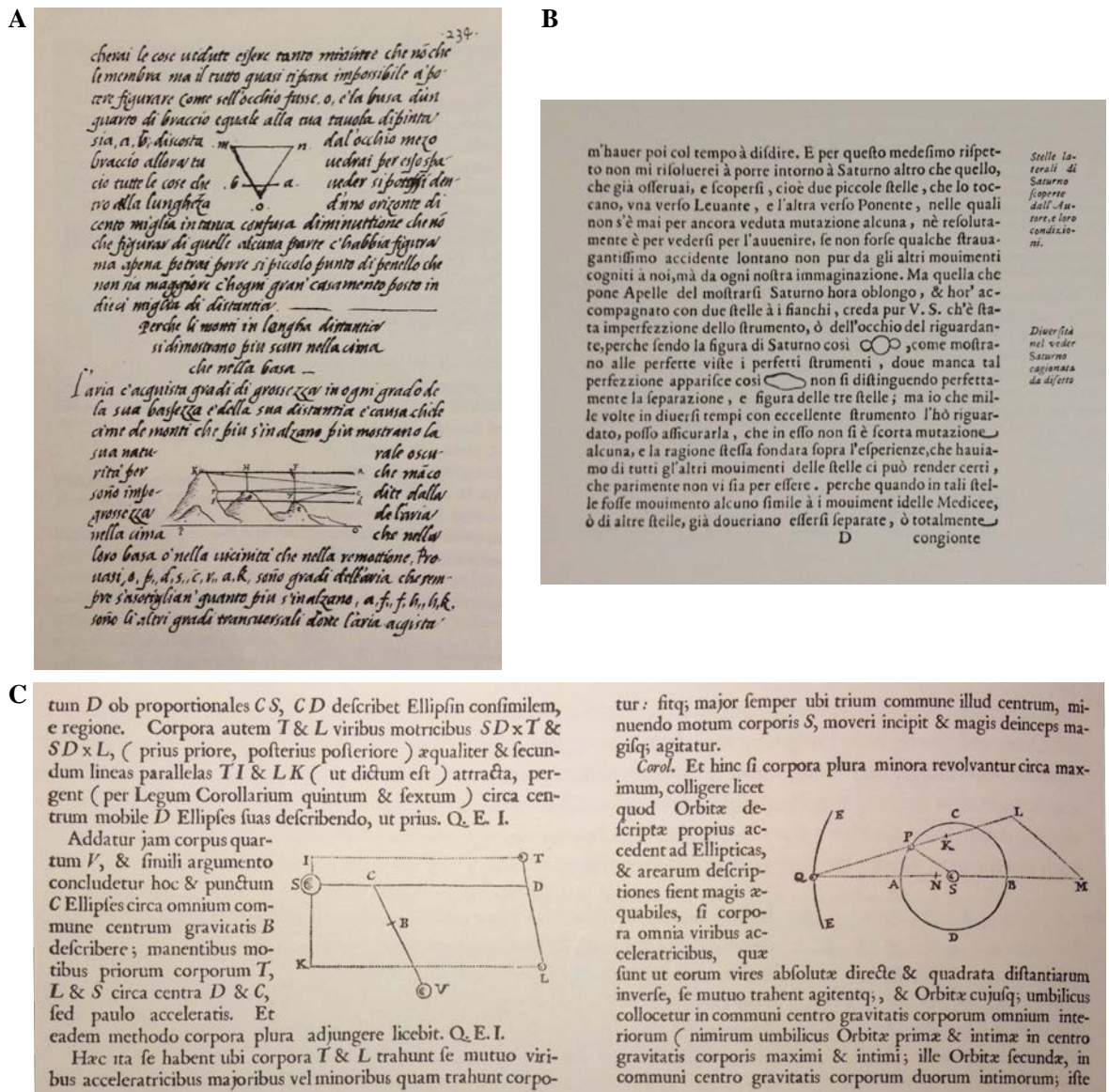
A

B

C

Figure C.9: *Pages from the manuscripts of Leonardo da Vinci, Galileo Galilei and Isaac Newton illustrating how they integrated figures with text.* (A) A page of a manuscript of Leonardo da Vinci demonstrating how he carefully placed figures within relevant text. *Source*: [Tufte, 1983, p. 182]; original source cited in Tufte's book: [da Vinci, 1956, p. 234 para. 827] (B) A page of a manuscript of Galileo Galilei in 1613 on his first telescopic observations of Saturn. Here Galileo uses drawings as sentence elements instead of words to describe the actual shape of Saturn (i.e., ⚭) and how it appeared in a telescopic view (i.e., ⬭). *Source*: [Tufte, 1990, p. 120]; original source cited in Tufte's book: [Galilei, 1613, p. 25] (C) An abstract for the book 'Principia' of Isaac Newton demonstrating how he carefully placed figures within relevant text. *Source*: [Tufte, 2006, p. 110]; original source cited by Tufte: [Newton, 1687, pp. 170,189]

Lessons can also be learned from books recording motion and dance choreographies. For instance, in Figure C.10A, multiple moves in a choreography are each described by a same structured column consisting of text describing the move, a floor plan with 3D figures carrying out the move, a floor plan with lines and symbols (it would have been helpful if a legend was provided) showing the movements in 2-dimensions, and the notes of the accompanying music. Actually, Figure C.10A illustrates the use of small multiples [Tufte, 1990, pp. 67-79], as the columns have the same structure to represent the same set of variables but are ordering based on another variable, in this case, the sequence of the moves.

Though many components are shown, the links between them are easily understood. The changes in the moves, as well as the timings, are easily understood due to the transparency and consistency in the display brought about by the use of small multiples, where each column describes ordered details (top to bottom) about a move and different columns describe the flow (left to right) between moves [Tufte, 2006, pp. 32-33]. In another example, Figure C.10B, the dance notations (describing the steps) flow along lines that indicate the movement the 3D figures laid just above them should carry out on the dance floor. The closeness of the dance notations to the lines and the lines to the figures clearly indicate the steps and the movement each dancer has to carry out without the need for further details [Tufte, 1990, pp. 114-15]. Tufte also notes that "the paper encoding reflects the refinement of the dance itself—a flowing and graceful line embellished by disciplined gesture" [Tufte, 1990, p. 114 para. 2] adding that some of the dance notations ([Hilton, 1981]—a review) are visually elegant irrespective of the illustrated moves. So, by looking at such displays, one can imagine the actual dance in real life.
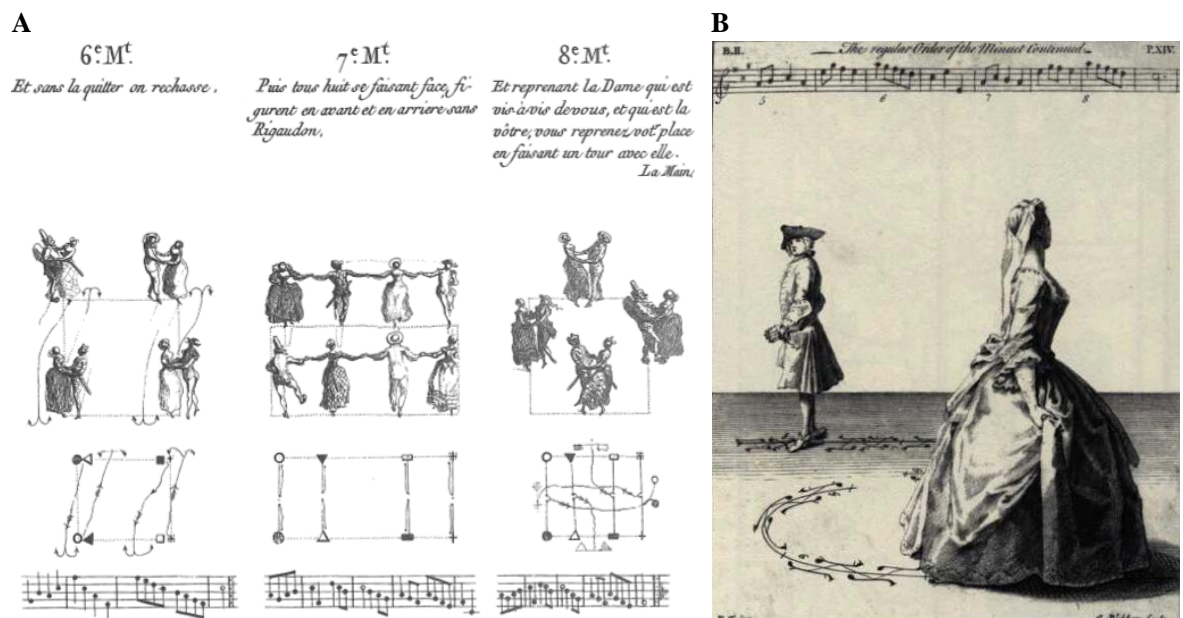

Figure C.10: *Recording dance choreographies.* (A) Three moves in a 'contredanse' choreography in a 1765 dance pamphlet [La Cuisse, 1762] which, as the title (i.e., 'Le Répertoire des bals ou théorie-pratique des contredanses, décrites d'une maniere aisée avec des figures démonstratives pour les pouvoir danser facilement, auxquelles on a ajouté les airs notés') indicates, figures and musical notations were intentionally included to easily demonstrate the moves and clearly relate them to the musical notes. *Source*: [Tufte, 1990, p. 33]; original source cited in Tufte's book: [La Cuisse, 1762] (B) A move in a minuet choreography in a 1735 book titled 'The Art of Dancing Explained by Reading and Figure' [Tomlinson, 1735]. *Source*: [Tufte, 1990, p. 114]; original source cited in Tufte's book: [Tomlinson, 1735]—Plate XIV

C.3.2 Tufte's View

Tufte [2001] notes that though different, text, graphics and tables are all presenting information and so, references such as "see Fig." should be avoided and instead, graphics and tables should flow into the text as part of the storyline in a way that they can use the same typeface and if some other text refers to the graphic or the table, then a replica is used and integrated with that specific piece of text. With the aim of avoiding the segregation of text, graphics and tables, he defined the principle of "data/text integration" (or the principle of "text/graphic/table integration"):

Data graphics are paragraphs about data and should be treated as such.

[Tufte, 1983, p. 181]

Tufte also encourages the integration of textual information (e.g., to explain the data or point out interesting observations and outliers), equations, legends and in some cases tables into the graphic to help the reader understand the graphic and its purpose in the storyline without having to revert back to the actual lengthy text [Tufte, 2001]. In fact, he claims that:

Words on and around graphics are highly effective—sometimes all too effective—in telling viewers how to allocate their attention to the various parts of the data display

[Tufte, 1983, p. 182]

This follows from studies in visual perception that suggest that when we look at a picture, textual instructions are decisive for our eye movements and thus, they determine the way we look at the picture [Gould, 1976].
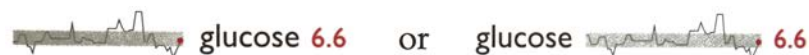
C.3.3 Sparklines

Tufte emphasized that, "Words and pictures belong together, genuinely together" [Tufte, 1990, p. 116] and as an example, he created visually distinct "wordlike graphics" known as 'sparklines' [Tufte, 2006, pp. 46-63]. So, instead of reporting the glucose level of a patient as

glucose 6.6

[Tufte, 2006, p. 47]

a sparkline could be added as in

glucose 6.6   or   glucose 6.6

[Tufte, 2006, p. 47]

to provide a context and better understanding of the relevance of the number without the need to check other information. In this case, the data-line depicts other previous glucose level results of the patient, the red dot relates to the current 6.6 measure indicating the data-line should be read from left to right (right-most value is the most recent), and the grey area illustrates the normal glucose range to find out whether the patient's levels have been within this normal range or not [Tufte, 2006, p. 47]. These sparklines could thus be embedded within the lengthy text, with Tufte defining them as

sparklines—small, high-resolution graphics usually embedded in a full context of words, number, images. Sparklines are *datawords*: data-intense, design-simple, word-sized graphics
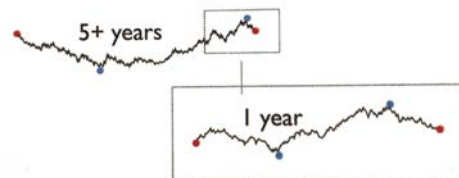
[Tufte, 2006, p. 47]

So similar to Galileo's method of integrating text and drawings, sparklines can be part of a sentence in the text or table. They also adopt the idea of small multiples (used widely include dance notation as we have seen earlier), so that, while within each sparkline the same data encoding and structure is used to represent the values of a set of variables (like letters in a word), different sparklines can be used to

represent different same-structured concepts whose details would be represented within the sparklines (like words in a sentence). In this way, once the data encoding mechanism is learned, any number of sparklines with the same structure can rapidly and effortlessly be analysed by just comparing the graphics. Like other displays using small multiples, sparklines also ensure data transparency and avoid confounding complex displays that combine the data into one complex graphic. So besides the glucose level, the doctors might want to check other variables and sparklines as in
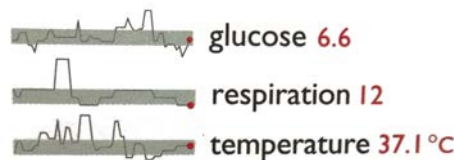
[Tufte, 2006, p. 47]

can facilitate analysis and comparisons. In other cases, a magnified view of a section of a sparkline can be illustrated using another sparkline as in

[Tufte, 2006, p. 50]

illustrating some measure of performance over time.

C.3.4 Ware's View

Similar to Tufte, Ware recommends (as one of his guidelines) to

> Place explanatory text as close as possible to the related parts of a diagram, and use a graphical linking method.
>
> [Ware, 2012, p. 333—G9.5]

to ensure cognitive efficiency, but claims that due to printing restrictions, this is not always possible. To justify his guideline [Ware, 2012, p. 333—G9.5], he points out a study [Chandler and Sweller, 1991] that demonstrated improved comprehension when text was integrated with the diagram (as in Figure C.11), in contrast to when the text and the diagram were segregated. The diagram represented an electrical system and the text provided instructions on how to run a test with this system. Consistent with previous work [Sweller et al., 1990, 1998], the authors argue that, since an integration of the text and the diagram was required for the subjects to understand and recall the test procedure, those with the text and the diagram segregated had to mentally integrate the two, leading to a cognitive overload due to the limited capacity of the working memory. In contrast, those with the text integrated with the diagram, did not have to retain much information in the working memory since the integration was already available. When, on the other hand, the text and the diagram did not necessarily have to be mentally integrated to be understood, no improvement was reported for text integrated with the diagram in comparison with text and diagram segregated. So it is likely that attention is devoted to just one of the representation (the text or the diagram) when the two are segregated. Thus, Sweller who devised the cognitive load theory advises that:

The cognitive effort required to mentally integrate disparate sources of information can be reduced or eliminated by physically integrating the various entities.

[Sweller, 1994, p. 302]

INSULATION RESISTANCE TESTS
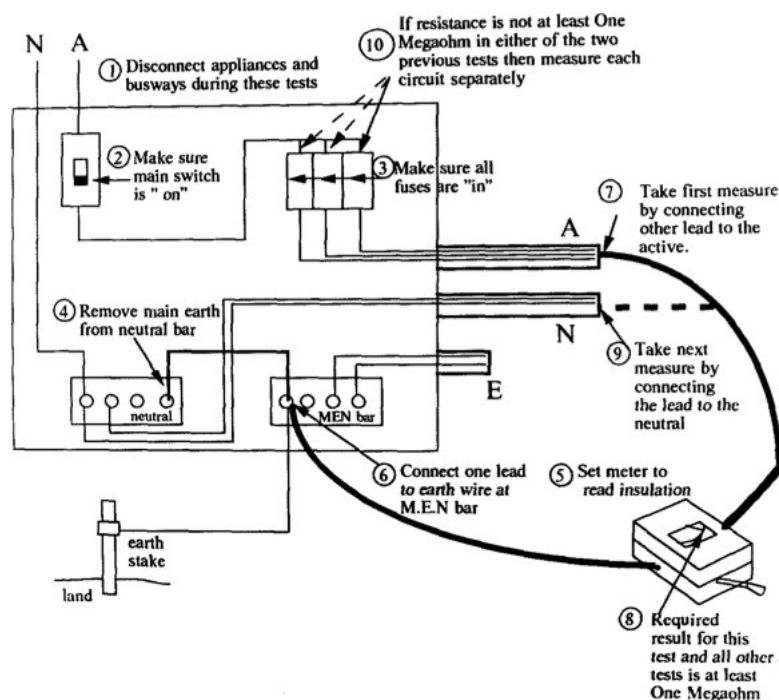
a) CONDUCTORS IN PERMANENT WIRING

Figure C.11: *Instructional text integrated with a diagram of an electrical system to illustrate how to run a test using the system.* In Chandler and Sweller's [Chandler and Sweller, 1991] first experiment, subjects using this visualization performed better than those using segregated text and diagram. *Source*: [Chandler and Sweller, 1991]—Figure 2

Ware argues that a narrative must direct and maintain the audience's attention and this is possible using language or visualization or both depending on what the narrative should communicate [Ware, 2008, pp. 129-45]:

Language can convey complex logical relationship between abstract ideas and support conditional actions. Visual media can support the perception of almost instantaneous scene gist, rapid explorations of spatial structure and relationships between objects, as well as emotions and motivations.

[Ware, 2008, p. 145]

He also ranked visualization techniques based on their cognitive strength to control the audience's attention, starting off from the most power techniques: movies and animated cartoons, strip cartoons, diagrams including pictures and photographs. Ware explains that the narrative of a diagram is not so

easily understood and readers "must actively discover the narrative in a diagram of still images" [Ware, 2008, p. 145]. Thus Tufte's emphasis to use words with graphics [Tufte, 2001] and a reason why textual instructions influence the way we look at pictures [Gould, 1976]. However, care is required to avoid unnecessary redundant text which could make the visualization cluttered and confusing [Tufte, 1990, 2001] and an example where this is particularly applicable is in diagrams in instruction manuals explaining how to assemble products for instance, furniture or appliances.

The cognitive process in using such assembling manuals has been studied and after experimentation using different representations, cognitive principles to design effective diagrammatic instructions for assembling furniture have been proposed, tested and integrated into an algorithm to automatically generate these instructions (e.g., Figure C.12) [Heiser et al., 2004]. Some fundamental principles included: ensuring transparency in the sequence of actions and consistency so readers can easily see the link between the frames (and similar to frames in a strip cartoon, one storyline is perceived); choosing an appropriate view point such that all parts are clearly visible and recognizable; showing the assembling actions rather than the final structure after the actions are carried out (in their experimentation, the authors demonstrated that diagrams showing actions rather than structural are easier to follow; such actions are visible in Figure C.12 where arrows and dotted lines are used to show direction and placement). Hence, in this case, words are not really necessary and most of information and actions can effectively be represented graphically using such principles. Ware [2008] notes that these assembling diagrams provide assistance for two different cognitive styles, namely, the narrative style where the reader carries out the sequence of steps as instructed and the exploratory style where the reader refers to different parts of the diagrams in no specific order and at any time to figure out the parts and how they should be assembled. If a video is used instead of the diagram, the exploratory style would not be supported as navigation through the sequence would not be practical.
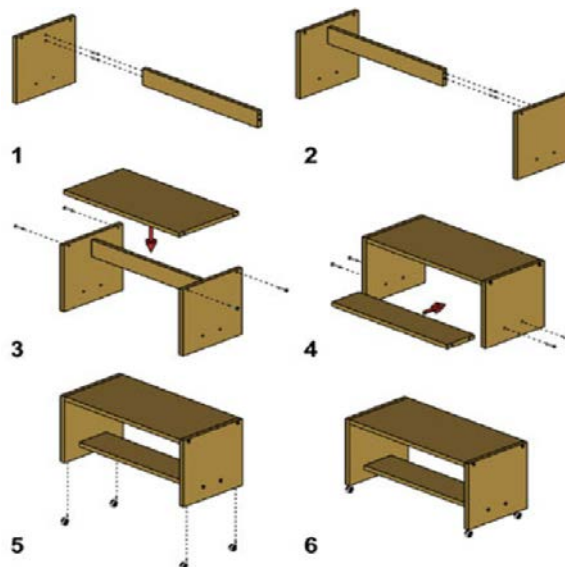

Figure C.12: *Automatically generated diagrammatic instructions to assemble furniture.* The cognitive design principles which Heiser et al. proposed and validated, have been integrated into an algorithm to automatically generate effective diagrammatic assembly instructions such as these. This diagram is an action diagram and not a structural diagram as it illustrates actions rather than the final structure at each step. *Source*: [Heiser et al., 2004]—Figure 7

C.3.5 In Journalism and Mass Communication

Visualizations are often used when reporting news events or stories. A study reported that articles with a visualization attract more readers and facilitate recall and comprehension, particularly when the visualization is large as readers read more of the story [Huh, 1993]. Other studies investigated how readers understand and use visualizations. For instance, a study [Pasternack and Utt, 1990] suggests that dominant graphics are often read before the title or text of the article, while smaller and less attractive graphics are read afterwards. However, the graphics were referred to mainly for content rather than just appearance. Some readers read the graphics first in preparation for the story in the text, others refer to them after reading the text as additional material, while other get all the information by just looking at the visualization. Thus, the authors argue that information should be provided in the visualization, but content in the article should not be repeated in the graphics.
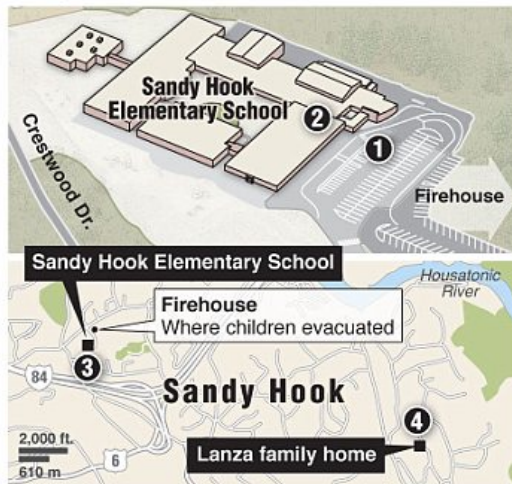
However, other studies in journalism and mass communication suggest that graphics that provide information that is also in the text aids readers' comprehension of the story. For instance, when reporting foreign news, a map showing the location of the incident [Griffin and Stevenson, 1992] as well as a graphic with background information about the referenced country [Griffin and Stevenson, 1994b], which information would also be provided within the story, help readers' understanding of the context of the event (some of these graphical features, including the map locator and a list of important facts in the story, are visible in Figure C.13). Also, the retention of statistical information related to a story is improved when the statistics are provided in both the text of the article and in a graphic [Griffin and Stevenson, 1996]. Visualizations known as 'how graphics' (e.g., Figure C.13) are often used in news articles to illustrate how events in the story unfolded. A study demonstrates that articles with both a 'how graphic' and a photo are the most effective in conveying stories followed by articles with just a 'how graphic' (when compared with articles with no visualizations or articles with just a photo and when the information provided by the graphics was also available in the text) [Stark and Hollander, 1990]. Another study recorded an improvement in recall when a 'how graphic' was provided with the text (again the information in graphic was also provided in the text) in contrast to text alone [Griffin and Stevenson, 1994a]. However, it was shown that though graphics that illustrate information that is already in the text facilitates understanding and improves recall, they detract the reader from information that is only provided in the text and not in the graphics [Ramaprasad, 1991]. Thus, providing the same information in both the graphics and the text is beneficial, but if the information cannot be provided in both, then it is best to place in the information in the text rather than the graphic.

Figure C.13: *A 'how graphic' published by the Daily Mail.* This "how graphic" illustrates how events unfolded during the mass shooting in Sandy Hook Elementary School in Newton, CT, USA on December 14, 2012. Some important facts about the story are displayed on the right. *Source*: [Bates and Peterson, 2012]

General reading behaviour assumptions have been outlined [Kress and van Leeuwen, 2006] and eye-tracking measurements were later used to investigate how many of these assumptions can be confirmed when reading a newspaper [Holsanova et al., 2006]. For instance, the eye-tracking measurements confirmed that readers often read the top of the page first and then the bottom, as it is often assumed that the most general information is at the top while the most detailed information is at the bottom [Holsanova et al., 2006]. Also, headlines and pictures are the first items that are looked at and items connected with lines, arrows or other framing mechanism are traced.

C.3.6 Storytelling

Often visualizations have a story to tell. They typically have a goal and an answer to a question in a particularly context, so that the context gives meaning to the presented data [Shapiro, 2004]. Thus, Shapiro [2004, p. 16] claims, "Question + Visual Data + Context = Story". The context must be clear and integrating text with the diagram can help. Example, the text at the top of Minard's [1869] flow map (Figure C.8) informs the reader what the presented data is all about and how the data was encoded.

Glassner [2004] focuses on the "bond" between the storyteller and the audience stating that the "essence of the bond between creator and audience is trust" [Glassner, 2004, p. 93] and the need to engage the reader. Both trust and engagement are important for information visualization. The audience has to trust the creator of the information visualization in that the presented information is correct, is in

the right context and is meant to inform rather than manipulate. The creator in turn must trust the reader to make appropriate use of the information and must ensure that the visualization is correct, appropriate and engaging for the information to be communicated effectively.

Engagement is particularly important when raising public aware of important issues. Hans Rosling, the co-founder of the award winning Gapminder [2013] for its visualizations, told the story (in around 4 minutes) of how wealth (based on earnings) and health (based on lifespan) in 200 countries changed over 200 years using both animated charts and spoken narrative to inform his audience about the relevant statistical data [Rosling, 2010]. Similarly, information designer Morelli managed to raise awareness of everyday water consumption from food (also known as virtual water, that is water that is used in the production of food as well as goods) through the use of visual displays, storytelling (in the form of written narrative) and interaction (allowing the readers to scroll through the story at their own pace), and which she refers to as an 'infographic story' [2012a] (accompanying webpage [Morelli, 2013]). Referring to how she got the public's attention when giving talks on the topic before she created the 'infographic story', Morelli stated:

> I discovered the power of words when they are in synch with strong images … I experienced the power of telling stories and I felt that this power does not lie simply in embedding diagrams or charts in a larger body of text but it lies in using a visual platform as a stage—a stage where words and images dance in sync in order to guide the audience through a journey.
>
> [Morelli, 2012b]

She also emphasized the need not to overload the audience with too much details and to make sure that at the end they get the main message. Following these ideas, she then created the 'infographic story' on virtual water [2012a], which is currently being supported by the United Nations Regional Information Centre for Western Europe (2013 being the International Year of Water Coorperation).

In some cases, using stories that people can relate to can help to effectively convey the message. Analogies have been used extensively both as a cognitive tool [Hesse, 1966; Holyoak and Thagard, 1995] and as a teaching aid (e.g., for statistics [Martin, 2003], mathematics [Novick and Holyoak, 1991] and science [Donnelly and McDaniel, 1993; Treagust et al., 1992]) with studies suggesting their effectiveness when, for instance, analogies based on prior knowledge are used to learn new material (e.g., [Schustack and Anderson, 1979; Wharton et al., 1994]), when new problems are solved by referring to previously solved similarly structured problems (e.g., [Quilici and Mayer, 1996; Renkl et al., 1998]), when the learner has little background knowledge about the material taught (e.g., when teaching science [Donnelly and McDaniel, 1993]), or when the analogy is used to aid learners develop their own knowledge of a new concept (e.g., [Treagust et al., 1992]).

However, an analogy without a visualization might not always be so helpful. No knowledge in calculus or climatology is needed to understand the link between global warming and increased carbon emission [Sterman and Sweeney, 2007] (i.e., since the rate of carbon emission in the atmosphere is faster than the drainage rate, carbon will accumulate and warm up Earth) as this issue is analogous to a bathtub with running tap water and an open drain [Archer, 2009; Sterman and Sweeney, 2007; Sterman, 2011]. Even so, highly educated people lack intuition about such dynamic systems be it the climate [Sterman and Sweeney, 2007] or simpler systems such as a bathtub and bank accounts [Cronin et al., 2009; Sterman, 2002; Sweeney and Sterman, 2000]. This is due to human cognitive limitations which impede us from understanding long term consequences and from reasoning about flows and

accumulations that change over time and space [Sterman, 1994, 2008, 2011]. Nonetheless, studies in various contexts such as the 'war on drugs' and automobile leasing [Sterman, 2000a, b, c] suggest that a picture of a bathtub with a tap and a drain can illustrate the presence of accumulations and clearly demonstrate the relationship between these accumulations and the flow [Sterman, 2010]. This would thus provide a visual analogy that is highly compatible with the verbal information and so, aid readers to process, understand and internalize the presented information [Ziemkiewicz and Kosara, 2008]. This led Sterman [2011] to propose pictures of bathtubs as a possible effective way to inform the public about the gravity of the issue and raise awareness of climate change. National geographic took up this concept and employed information designer Nigel Holmes [2013] (well-known for his graphic designs to convey information to laymen) to create a graphic of this bathtub analogy and tell the story about the relationship of carbon emission and global warning (Figure C.14).
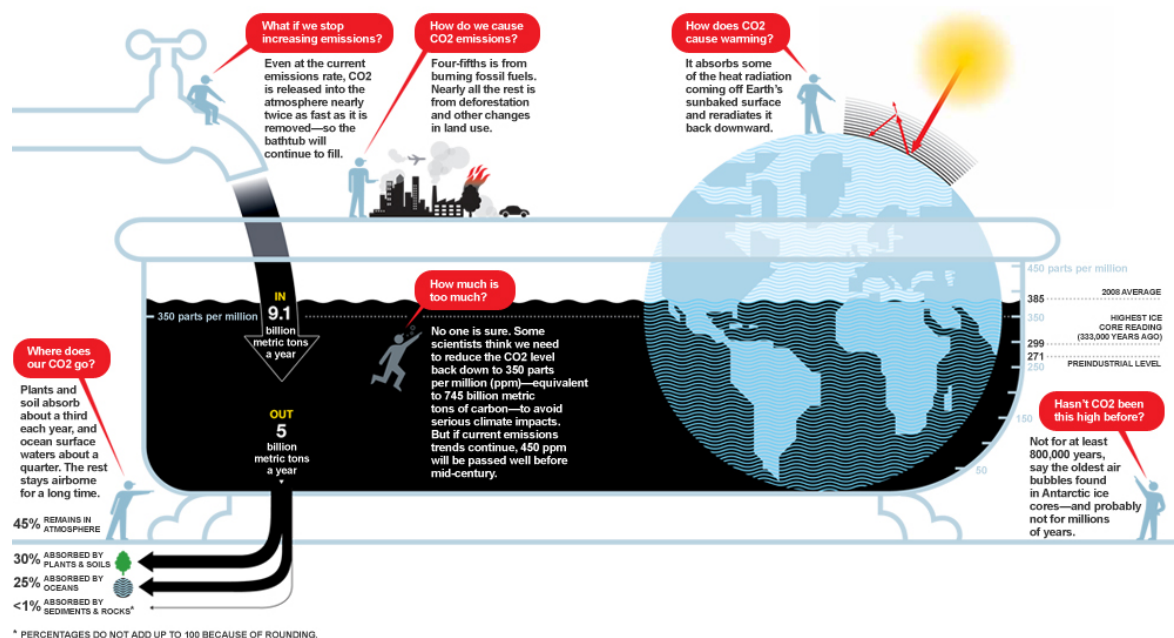

Figure C.14: *The Carbon Bathtub published by the National Geographic Magazine to explain global warming due to carbon emission.* "It's simple, really: As long as we pour CO_2 into the atmosphere faster than nature drains it out, the planet warms. And that extra carbon takes a long time to drain out of the tub." [National Geographic, 2009]. Created by Holmes and the Global Carbon Project, to convey scientific facts to the public and to raise awareness of the gravity of the problem. *Source*: [National Geographic, 2009]

Morelli used a variety of analogies when designing the illustrations of Tony Allan's book 'Virtual Water' [Allan, 2011]. Tony Allan is well-known for his expertise in sustainable water development and his contributions in communicating related issues to the general public, and this book was another of the latter contributions. For instance, Morelli used water filled bathtubs to demonstrate the litres of water that are consumed daily by a non-vegetarian (Figure C.15). The readers would have definitely seen a bathtub before and thus, they can relate to it and better understand the quantities of water the author is talking about.
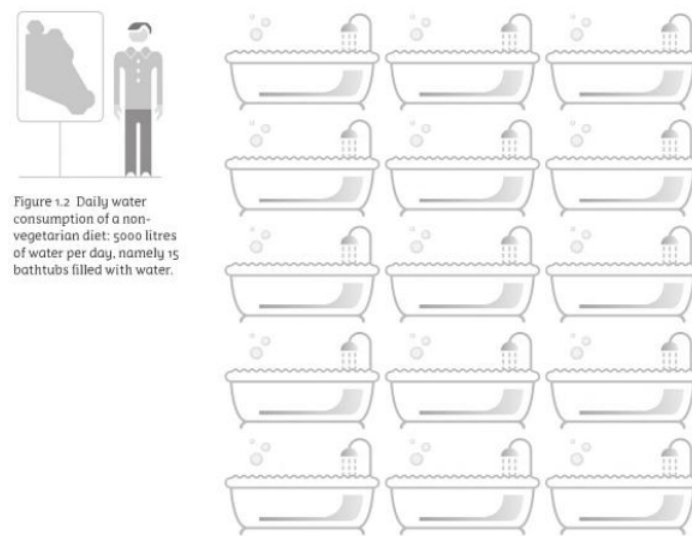
Figure 1.2 Daily water consumption of a non-vegetarian diet: 5000 litres of water per day, namely 15 bathtubs filled with water.

Figure C.15: *Using water filled bathtubs to help readers understand how much 5,000 litres of water is in Tony Allan's book, 'Virtual Water' [2011].* This graphic was designed by Angela Morelli to demonstrate that 5,000 litres of water consumed in a non-vegetarian diet daily is equivalent to 15 water filled bathtubs. It might be hard to understand how much 5,000 litres of water is, but not in the context of bathtubs which all readers have seen before. *Source*: [Allan, 2011]—Figure 1.2

A story can express

> great quantities of information in relatively few words in a format that is easily assimilated by the listener or viewer. People usually find it easier to understand information integrated into stories than information spelled out in serial lists (such as bulleted items in an overhead slide). Stories are also just more compelling.
>
> [Gershon and Page, 2001]

In fact, they are used in various contexts (including, for instance, in organizations to transfer knowledge [Dennehy, 2001; Sole and Wilson, 2002]). After all, "Knowledge is Stories" [Schank, 1990, p. 1]. However, similar to Ware's argument about narrative (Appendix C.3.4), both visualizations and language can be used to communicate a message and a story. Visualizations are more engaging than text [Gershon and Page, 2001] and in cases such as instruction manuals showing the required actions to, for instance, assemble a furniture, pictures with just numbers could be enough (e.g., Figure C.12). But in most other cases, visualizations alone cannot tell all the story, particularly since the context must be clear for the visualization to have a meaning [Shapiro, 2004]. So, it is sensible to tell the story by using both text and graphics [Gershon and Page, 2001]. However, since information visualization visualizes abstract concepts, appropriate analogies and "visual metaphors", such as the bathtub used in both National Geographic's graphic on global warming (Figure C.14) and similar bathtubs used by Morelli to illustrate quantities of water (Figure C.15), should be identified and used [Gershon and Eick, 1995; Gershon and Page, 2001, p. 33] (this has also been studied for advertising [Forceville, 2002]).

To convey a story using visualizations, a number of storytelling techniques should be considered [Gershon and Page, 2001] (see also [Gershon and Eick, 1995; Pinker, 1994; Schank, 1990]), such as: clearly illustrating the big picture, the context and the setting; providing a sense of continuity between different visualizations in the story-line; providing magnified views to avoid ambiguities; using redundancy effectively to help the reader link text to the pictures [Gershon and Page, 2001]. Some of

these storytelling techniques actually overlap with previously discussed principles such as Tufte's 'text/graphic/table integration' principle [Tufte, 2001] (Appendix C.3.2), the cognitive priniciples defined to create effective instruction manuals to assemble goods [Heiser et al., 2004] (Appendix C.3.4), and the techniques used in journalism and mass communication (Appendix C.3.5).

Segel and Heer [2010] provide a general and thorough review of how stories can be related through information visualization. They argue that the readers do not get to know the story from just the details provided by the authors of the visualization, but more than that the readers often attempt to actively explore and discover more about the story. Thus, the need for interactive visualizations to facilitate this exploration. Segel and Heer analysed and classified the design features of various visualizations that relate a story in areas such as budget predictions, reporting of war events and discovery of trends in human development (e.g., Gapminder's [2013] interactive visualizations), and defined a framework of design strategies that creates effective storytelling visualizations. However, they conclude that "there is much that remains to be understood" [Segel and Heer, 2010].

Designing engaging visualizations that are able to attract the reader's attention and effectively convey information to a wide, diverse audience in a concise but informative manner is challenging. It requires designers to blend "the skills of computer science, statistics, artistic design and storytelling" [Cukier, 2010] to provide "readers material that is worthy of their scan that makes them stop scanning and start reading." [Garcia and Stark, 1991].

C.4 Euler-Like Diagrams with Glyphs

Closed curves are sometimes drawn around elements in a diagram to visualize the data set membership of elements and relationships between the data sets. Such a visualization would be similar to an Euler diagram with glyphs.

Different irregular curves are used to visualize the groups including rounded rectangles as in ComED and DupED [Riche and Dwyer, 2010] (e.g., Figure C.16A), smooth irregular curves as in Set Visualiser [Wyatt et al., 2009; Wyatt, 2010] (e.g., Figure C.16B) or non-smooth irregular curves [Simonetto et al., 2009] (e.g., Figure C.16C).
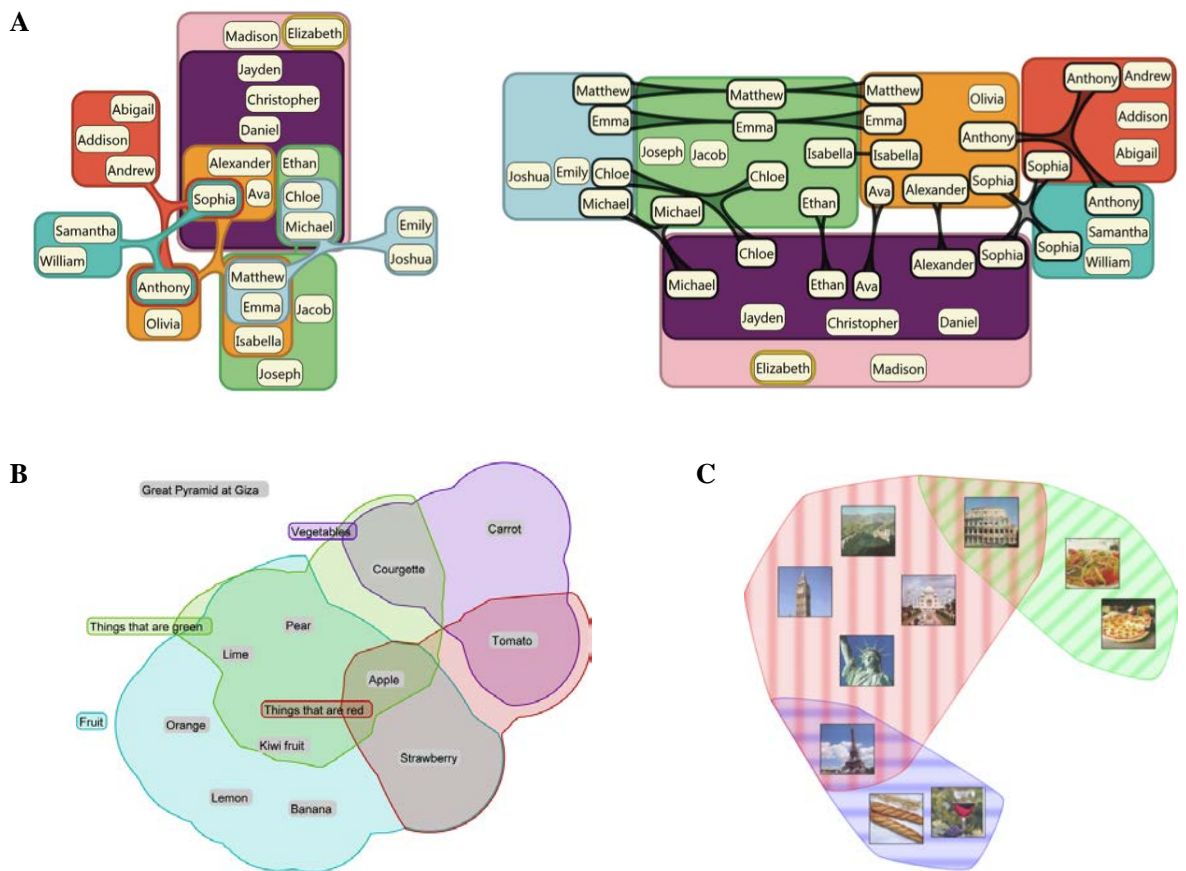
A

B

C

Figure C.16: *Examples of not area-proportional Euler-like diagrams with glyphs to represent data set membership of elements and their relationships.* (A) Generated using ComED (left) and DupED (right). The diagram represents relationships in a social network. *Source*: [Riche and Dwyer, 2010]—Figure 1. (B) Generated using Set Visualiser. The diagram represents relationships between 'Fruit' (blue curve), 'Things that are green' (green curve), 'Vegetables' (purple curve) and'Things that are red' (red curve) and notes disjointness of these data set with the set 'Great Pyramid at Giza' (grey label on the left but no curve). *Source*: [Wyatt, 2010]. (C) Generated using Simonetto et al.'s technique. The diagram represents relationships between world monuments (red curve), Italian culture (green curve) and French culture (blue curve). *Source*: [Simonetto et al., 2009]—Figure 1.