REVEALING THE SCIENTIFIC BASIS OF GRAPHICAL REPRESENTATION DESIGN

Aantal woorden/ Word count: 21,621

Sarah Pissierssens
Stamnummer/ Student number: 01200213

Promotor/ Supervisor: Prof. dr. Geert Poels
Commissioner/ Commisaris: Prof. dr. Frederik Gailly

Masterproef voorgedragen tot het bekomen van de graad van:
Master's Dissertation submitted to obtain the degree of:

Master of Science in Business Engineering

Academiejaar/ Academic year: 2016 - 2017
REVEALING THE SCIENTIFIC BASIS OF GRAPHICAL REPRESENTATION DESIGN

Sarah Pissierssens
Stamnummer/ Student number : 01200213
Promotor/ Supervisor: Prof. dr. Geert Poels
Commissioner/ Commisaris: Prof. dr. Frederik Gailly

Masterproef voorgedragen tot het bekomen van de graad van:
Master's Dissertation submitted to obtain the degree of:
Master of Science in Business Engineering

Academiejaar/ Academic year: 2016 - 2017
Permission

I declare that the content of this Master’s Dissertation can be consulted and/or reproduced if the sources are mentioned.

Name student: Sarah Pissierssens

Signature:
Summary

From the start, graphical representations have played a central role in *Information Systems & Computer Science*. While quality frameworks for visual representation semantics have already matured, design rationale for the visual syntax of graphical representations has been largely absent. This paper aims to provide an initial scientific basis for graphical representation design and presents 12 principles for qualitative graphical representation design. Three types of design principles are distinguished: fundamental, practical and integrating principles. The fundamental principles form an explanatory backbone for the practical principles, which direct the diagram engineer through the design process. Last, the integrating principles aim to resolve conflicts between principles and balance diagram requirements. The 12 graphical representation design principles were formulated by bringing descriptive theory from fields like *Cognitive Sciences* together with prescriptive design guidelines and rules from practice. The principles can be used for designing, comparing, evaluating and improving graphical representations.

As a secondary effect, this paper aims to cultivate a self-conscious design culture and contribute to a scientific basis for graphical representation design.
Acknowledgements

When I decided on the topic of my master thesis, I knew I was making a pivotal decision. This topic was going to become the centre of the challenging ‘thesis experience’ I had heard everyone talk about. I purposefully looked for a stimulating subject which was a bit out of my comfort zone and not too related to my Major in Operations Management. I also was looking for an approachable promotor who would coach me rather than firmly direct me.

I am thankful to Geert Poels, professor at Ghent University, for offering me both. While the topic was set onto research in the field of quality in conceptual modeling with a focus on user-perspective, he left some blanc spaces for me to fill in according to my own findings. I want to thank him for his support and his trust when I altered the direction of my thesis quite significantly. Thank you for the opportunity to write my master thesis on such an interesting and upcoming topic, and for all the helpful comments and thoughtful remarks you gave.

A second person I want to greatly thank is Jan Claes, postdoctoral researcher at Ghent University. He has been a great advisor and companion in this academic adventure, guiding me through important content-related turning points, but also giving me some invaluable practical tips and best practices of his own. Jan, thank you a lot for all your energy, the trust and the motivational talks, my thesis would definitely not have been the same without your support and help!

Last but not least, I would like to thank my roommate Anouk for making me amazing coffee, my mother for cooking me delicious exotic food to keep me energised, my father for putting small things into perspective, and finally, Guillaume for being my greatest supporter and trustee. All of the little things have made this grand journey so much more successful, memorable and enjoyable!

Sarah Pissierssens
Gent, May 2017
# Table of Contents

**Permission** ......................................................................................................................................... i
**Summary** ........................................................................................................................................... i
**Acknowledgements** ............................................................................................................................ ii

**Table of Contents** ............................................................................................................................... iii

**List of acronyms** ..................................................................................................................................... v

**List of figures** ......................................................................................................................................... vi

**List of Tables** .......................................................................................................................................... vii

1. **Introduction** ...................................................................................................................................... 1
   1.1. Importance of graph representations in Information Systems & Computer Science .................. 1
   1.2. Conceptualising graphical representation design ................................................................. 2
       1.2.1. Graphical representation design: a definition ............................................................ 2
       1.2.2. Graphical versus textual representations ..................................................................... 3
       1.2.3. The anatomy of graphical representations ....................................................................... 3
       1.2.4. Quality in graphical representations: the dependent variable of graphical representation design ........................................................................................................................................ 4
   1.3. The mole in graphical representation design: visual dialects .................................................... 6
   1.4. Scope and objective of this paper ............................................................................................... 7
       1.4.1. Research question ............................................................................................................... 7
       1.4.2. Research objectives ........................................................................................................... 7
       1.4.3. Scope .................................................................................................................................. 7

2. **Methodology** ..................................................................................................................................... 9
   2.1. Revealing the missing link: a holistic literature study of reference domains ......................... 9
   2.2. Formulating the missing link: the graphical representation design principles ..................... 9

3. **Design principles for graphical representations** ............................................................................. 11
   3.1. A scientific basis for graphical representation design ............................................................. 11
       3.1.1. Reference Domains ........................................................................................................... 12
       3.1.2. Emerging graphical representation design principles .................................................. 12
   3.2. The 12 principles of graphical representation design ............................................................ 12
       3.2.1. Fundamental Principles .................................................................................................. 13
       3.2.2. Practical Principles ......................................................................................................... 17
       3.2.3. Integrating Principles ..................................................................................................... 40

4. **Conclusion and discussion** ............................................................................................................... 47
   4.1. The physics of graphical representations: a theory for graphical representation design ....... 47
4.1.1. The Fundamental Principles ................................................................. 47
4.1.2. The Practical Principles ........................................................................ 47
4.1.3. The Integrating Principles ................................................................. 47
4.2. Practical evaluation and significance ...................................................... 47
4.3. Theoretical evaluation and significance .................................................. 48
  4.3.1. Prior research and current problem situation ...................................... 48
  4.3.2. Scoping this paper on the graphical representation anatomy ............ 49
4.4. Limitations and further research ............................................................ 49
4.5. Wider significance ................................................................................... 49
References .................................................................................................... 51
Appendix A ................................................................................................. 63
Appendix B ................................................................................................. 65
Appendix C ................................................................................................. 66
Appendix D ................................................................................................. 67
Appendix E ................................................................................................. 68
Appendix F ................................................................................................. 69
List of acronyms

BPM.............................................................Business Process Models
SE.............................................................System/Software Engineering
UML.............................................................Unified Modelling Language
List of figures

Figure 1: While The Physics of Notation has focussed exclusively on the top-left hand quadrant, this paper will focus on the bottom-left quadrant of the figure. Source: Moody (2009) .......................................................................................................................... 8

Figure 2: Shannon & Weaver’s Communication Theory applied to graphical representations. ........................................................................................................................................................................... 11

Figure 3: Reference domain matrix .......................................................................................................................... 11

Figure 4: Illustration of the Gestalt Law of Proximity .......................................................................................... 18

Figure 5: Illustration of the Gestalt Law of similarity .......................................................................................... 19

Figure 6: Illustration of the Gestalt Law of Symmetry .......................................................................................... 19

Figure 7: Illustration: fading at branch level. Source: Ding & Mateti (1990) .................................................. 23

Figure 8: Illustration: fading at abstraction-level. Source: Ding & Mateti (1990) ........................................... 23

Figure 9: Illustration of the Principle of Good Continuity. Path a-b is easier to follow than path c-b. Source: Field et al. (1993) ............................................................................................................... 25

Figure 10: Illustration of the importance of direction of information flow for the understandability of diagrams. Source: Ding & Mateti (1990) ................................................................. 25

Figure 11: Illustration of a cognitive concept map. Source: Siau & Tan (2005) .................................................. 28

Figure 12: Illustration of a context diagram. Source: Berenbach (2004) ......................................................... 29

Figure 13: Illustration of a summary note of a UML diagram. Source: Amber (2005) ................................. 30

Figure 14: Illustration of stretch text: the use of OCL on a UML diagram (indicated in red). Source: Ambler (2005) .................................................................................................................................................. 30

Figure 15: Human processing of graphical representations. Source: Moody (2009) ........................................... 31

Figure 16: Illustration of depicting a crossing of lines as a jump. Source: Ambler (2005) ............................ 32

Figure 17: Illustration of the minimum angle principle: the edges on the right hand side (90°) are better discriminable than the left hand side (acute angle). Source: Ware et al. (2002) .... 32

Figure 18: Illustration of conflicting rules. (a) satisfies the rule of ‘minimise number of crossings’, while (b) satisfies the rule of ‘minimise total path length’. Source: Ding & Mateti (1990) .................................................................................................................. 44

Figure 19: Illustration of the Gestalt law of good continuation. The continuity of the branches is easily identified although the grid disrupts the image. Source: Field et al. (1993) .......... 65

Figure 20: Comparing graphical representations on the Behavior Characteristics Spectrum: from FFBDs (Function Flow Block Diagrams) to DFDs (Data Flow Diagrams). Source: Long (2002) .................................................................................................................................................. 66

Figure 21: Illustration of spatially integrated text and picture. Spatial proximity did not improve learning any further when text was segmented and pictures were labelled. Source: Florax & Ploetzner (2010) .................................................................................................................. 67

Figure 22: Illustration of learning material with segmented text and labelled picture. Source: Florax & Ploetzner (2010) .................................................................................................................................................. 67
List of Tables

Table 1: Modularisation Principle: Overview................................................................. 21
Table 2: The different visual variables: power = scale of measure, capacity = maximal number of instances to be used on a diagram. Source: Moody (2009).............................. 36
Table 3: Metrics for graph drawing (Purchase, 2002)................................................... 45
Table 4: The retinal variables apply for points, lines and areas. Source: Roberts (2000)..... 68
Table 5: Source: Illustration of the application of the seven aesthetic metrics of Purchase. Purchase (2002) ......................................................................................................................... 69
List of definitions

Apprehension principle:
The Apprehension Principle states that graphical representations should be readily and accurately perceived and conceived. The principle argues that diagram elements which will not be accurately or easily comprehended, should be left out. For example, in the case of a road map, adding the precise angels of turns and width of roads will add unnecessary complexity to the diagram and should be omitted (Tversky, Morrison, & Betrancourt, 2002).

Cognitive fit theory:
Cognitive Fit Theory (CFT) states that user performance is improved when task material representation matches with the task to be executed. For example, presenting statistical data set in a list or table format facilitates the calculation of statistical averages of these series of numbers (Vessey & Galletta, 1991). Additionally, CFT argues that a match between the user and end-task also has a positive effect on performance, meaning that a graphical or logical oriented user will work more efficiently on respectively the lay-out of a model, or the semantical correctness of a model.

Cognitive load theory:
Human cognitive processing is limited. Cognitive processing takes place in the working memory (or short-term memory), which is restricted in the number of elements it can simultaneously hold. Miller (1956) argued that working memory can only accommodate 5 plus/minus 2 informational elements at a time. Recent work finds that working capacity is actually limited to 3 to 4 elements at a time (Cowan, 2010; Van Merriënboer & Sweller, 2005). In this context, Cognitive Load Theory (CLT) states that when working memory is overloaded (i.e. more than 3 to 4 elements need to be processed simultaneously), learners make more errors, learn more slowly and little room is left for learning (Sweller, 1988). Additionally, CLT suggests that instructional design of learning material can decrease the cognitive load on the working memory and therefore improve learning.

Coherence effect:
Information transfer and learning is improved when extraneous or redundant elements are excluded from the learning material (Mayer & Moreno, 2003).

Component-fluency principle:
Training routine aspects, or, consistent components of a task up to a very high level of automaticity, in addition to training the whole task, has a positive effect on learning (in
particular strengthening cognitive schemas) and transfer of the whole task. (Merrienboer & Kester 2005)

Computational offloading:
Computational offloading refers to the ability of diverse external representations reduce the amount of cognitive effort required to solve informationally equivalent problems (Larkin and Simon, 1987). When a graphical representation makes a presented concept significantly more easy to grasp, computational offloading is high.

Congruence principle:
The congruence guideline argues that format and structure of graphical representations should correspond to the content they represent (Tversky, Morrison, & Betancourt, 2002). For example, a graphical representation of to represent the hierarchical structure of an organisation will be effectively represented by a tree diagram.

Dual Channel Theory
Mayer & Moreno (Moreno & Mayer, 1999) state that humans possess separate channels for processing pictorial and verbal information. The dual channel processing theory is based on an initial work by Paivio (1986), who found that humans have distinct channels for processing visual and auditory information.

Dual Coding Theory
Building on Dual Channel Theory, Mayer & Sims (1994) argues that through processing verbal and pictorial information in two separate channels, both a word-based mental representation and a picture-based mental representation is built in long-term memory. When referential connections are made between the two mental representations, information recall is improved, since the human mind can now access the knowledge in multiple ways.

Elaborative learning
Elaborative learning describes the process of learning concepts with elaborative (i.e. extensive and detailed) supporting information (Van Merriënboer & Kester, 2005).

Multimedia effect:
Using Dual Coding Theory as a backbone, the multimedia effect states that students learn more deeply from words and pictures than from words alone (Moreno & Mayer, 1999). Therefore, the use of both verbal and graphical representation formats is promoted to improve information conveyance.
Repleteness effect:
The repleteness effect states that graphical representations which do not use visual notations which were previously acquired by the user, do not improve learning and do not lead to computational offloading effects (Westelinck, Valcke, De Craene, & Kirschner, 2005).

Spatial contiguity effect:
Spatial contiguity effect states that learning is improved when text is placed in close spatial proximity to corresponding pictures (Moreno & Mayer, 1999).

Split-attention effect:
When content elements are separated either in space or time, learners must invest additional cognitive resources mentally to integrate these elements in order to understand the materials (Chandler & Sweller, 1992). When related elements are split over space, extra cognitive load is imposed by searching and matching elements. When associated elements are split over time, effort from the learner is required to recall existing knowledge and integrate it with the displayed content. Reducing or eliminating the need for mental integration of split content sources, makes that cognitive resources can instead be directed towards learning.
1. Introduction

1.1. Importance of graph representations in Information Systems & Computer Science

The increasingly data-driven and complex problem domains with which today’s organisations are confronted, have led to growing attention towards graphical representations. Data is often multivariate and data sets are becoming notoriously big (Ellis & Dix, 2007; Koshman, 2010). When data becomes more abstract and grows in size, it easily becomes incomprehensible (Buagajska, 2001). As a result, people are looking at graphical representations to enable reduction of data complexity\(^1\) to facilitate data exploration and to ease sense-making (Ellis & Dix, 2007).

The increasing interest in graphical representation design in *Information Systems and Computer Science* is reflected in the growing research about visualisation techniques (Moody, 2009) and quality frameworks for graphical representations (Krogstie, Sindre, & Jørgensen, 2006; Nelson, Poels, Genero, & Piattini, 2012). Studies frequently refer to research in *Psychology* that shows that textual representation and numbers are often insufficient for efficient complex problem solving, and that graphical representation is desirable (Gaissmaier et al., 2012; Regnell, Andersson, & Bergstrand, 1996). Qualitative graphical representations have been found:

- to decrease ambiguity (Regnell et al., 1996),
- to aid processing (Stenning & Oberlander, 1995),
- to promote completeness of information through inference-making (Cox, 1999; Larkin & Simon, 1987) and through gap analysis (Novak & Cañas, 2008)
- to support computational processes by making information search more efficient (Larkin & Simon, 1987)
- to facilitate mental integration of multiple information sources (Sweller, Merriënboer, & Paas, 1998),
- to improve understanding (Gaissmaier et al., 2012), learning (Novak & Cañas, 2008; Prezler, 2004) and memorisation of knowledge (Tversky, 2001), and

\(^1\) We define 'complex information' as information high in abstraction and element interactivity, and becoming unintelligible in isolation (Ginns, 2006). Data complexity gets aggravated when the amount of information grows.
to support the creation, articulation, communication and archiving of knowledge (Novak & Cañas, 2008).

From the start, graphical representations have been central to the complex design activities in software and system engineering (Hjalmarsson & Lind, 2004; Krogstie et al., 2006; Long, 2002). Today, the emergence of Model Driven Architecture (MDA) is making conceptual models and visual notations for system engineering, such as UML and Archimate, even more relevant by promoting their use throughout all the phases of the system engineering project (Wagelaar & Van Der Straeten, 2007). Next to supporting engineering activities, graphical representations also have a pivotal role in the communication of complex information to a wide range of stakeholders. In an organisational context, process models and work flow diagrams are used for process and information management at different management layers (Nelson et al., 2012). To become a ‘knowledge creating company’, Nonaka (2007) underlines the importance of capturing the tacit knowledge of experts, for example in graphical representations. This can be achieved through techniques such as cognitive mapping. Next, in the context of Environmental and Earth Sciences, visualisation is a critical tool in the modelling of scientific environmental metadata, supporting scientific environmental discovery, and monitoring global environment change (Koshman, 2010; Zahid, Wanger, & Kochevar, 1994). The same holds for the conveyance and understanding of health-related statistical information, where graphical information can decrease life-threatening ineffectiveness compared to numerical information (Gaissmaier et al., 2012). Finally, in the transport industries, such as the railway industry, visualisation techniques ensure that a variety of end-users are able to extract the right conclusions from safety and risk analysis models (Figueres-Esteban, Hughes, & Van Gulijk, 2015).

It is thus clear that graphical representations are highly relevant across many fields, including Information Systems and Computer Science, but how do we optimise their quality to make them maximally effective? This leads us to the question of what a graphical representation really is and what quality in graphical design entails.

1.2. Conceptualising graphical representation design

1.2.1. Graphical representation design: a definition

A graphical representation is the product of making abstraction of some of the real-world complexity (Nelson et al., 2012; Rockwell & Bajaj, 2005) by purposefully representing selected information objects and their interrelationships, in the context of a specific design goal and audience (Cox, 1999).
Typically, graphical representation design merges facets of aesthetics, information flow and interactivity (Lau & Moere, 2007), and involves content picking, showing relationships within that content, visualising the information and lastly, timely presenting the information elements. We will refer to this later as ‘the 4 dimensions of graphical representation design’.

1.2.2. Graphical versus textual representations

Diagrammatic representations differ from textual representations on two levels: the encoding and the decoding level (Moody, 2009). To start, in contrast to one-dimensional textual (sentential) representations, diagrams\(^2\) are encoded in a two-dimensional solution space (Larkin & Simon, 1987). Secondly, according to dual channel theory (Mayer & Moreno, 2003), humans decode and process visual and textual information in separated channels. Consequently, different design principles are required for building textual or visual representations (Moody, 2009).

1.2.3. The anatomy of graphical representations

In building a visual representation, two interconnected ingredients are optimised: the visual language (or visual notation) and the diagrammatic spatial arrangement. The first ingredient, visual notation, is composed out of a combination of visual vocabulary (graphical symbols), visual grammar (compositional rules) and visual semantics (the meaning of the graphical symbols) (Buagajska, 2001; Moody, 2009). The second ingredient, diagrammatic spatial arrangement, reflects the spatial properties of the diagram like spatial proximity, structural arrangements, sequencing, layering, positioning, direction of information flows and spatial density (Buagajska, 2001).

However, not any combination and choice in visual variables will result in a qualitative graphical representation (Austin, 2009). Research has shown that only carefully designed diagrams are advantageous for representing complex information (Tversky, Morrison, & Betrancourt, 2002). Even more, unthoughtful visual design can negatively impact learning for both high and low prior knowledge learners (Schnotz & Bannert, 2003) by increasing cognitive load and causing split attention effects (Austin, 2009). It is thus very important to adhere to proper design rationale when designing a diagram, in order for it to be effective. This leads us to the question of what proper design rationale is and what a good, qualitative graphical representation is made of.

\(^2\) In this paper, we will also refer to graphical representations as ‘diagrams’ or ‘visual representations’.
1.2.4. Quality in graphical representations: the dependent variable of graphical representation design

Today, a major problem in graphical representation design is the absence of a clear and overarching design goal (a dependent variable) for designers to work to (Moody, 2009). The absence of this dependent variable is related to the lack of a centrally stated definition and common understanding of quality in graphical representations (Berenbach, 2004; Rockwell & Bajaj, 2005). What is not concisely and clearly defined, is hard to measure or improve (Ghylin et al., 2007), which might be the reason why the notion of quality in conceptual modelling is still such an immature and rapidly evolving concept (Nelson et al., 2012).

A. Fit for purpose

In the existing literature, many approaches to the concept of quality are taken. Garvin (1984) distinguishes three traditional approaches: quality as innate excellence (Tuchman, 1980), quality as requirements conformance (Crosby, 1979) and finally, quality as fitness for use (Juran & Godfrey, 1998). Because of the user-oriented nature of graphical representations (Moody, 2009), we adhere to the latter approach and thus we define quality of graphical representations in terms of their fit-for-purpose. This is in line with various other studies that perform user-goals analysis to determine diagram effectiveness (Casner, 1989; Cox, 1999; Novak & Cañas, 2008; Roth & Mattis, 1990).

Concretely, defining quality in graphical representations as fit-for-purpose means that a diagram needs to be suitable for achieving its goals and it needs to be ready to use. In other words, a graphical representation cannot be of high quality when it doesn’t answer some particular question we seek to answer (Novak & Cañas, 2008). This leads us to the question: what purposes do visualisations serve in general?

B. Purposes of graphical representations

In literature, we find a wide variety of proposed largely/mainly domain-independent-purposes for graphical representations:

- to facilitate the lookup, n-wise comparison and pattern-marking of data (Roth & Mattis, 1990),

- to serve as a diagnostic tool (e.g., for cognitive mapping to determine knowledge gaps (Fiol & Huff, 1992; Siau & Tan, 2005)), to identify bottlenecks, defects and design flows (Sadowska, 2013)),
- to show functional relations between object elements (Chen, 2005; Roth & Mattis, 1990),

- to display data distributions (Roth & Mattis, 1990),

- to coordinate different aspects of real-life task performance (Van Merriënboer & Kester, 2005),

- to support software-development (Rockwell & Bajaj, 2005; Sadowska, 2013),

- to serve as an informal language to improve understanding for novices (Fiol & Huff, 1992; Siau & Tan, 2005),

- to improve, adapt, understand, visualise, automate and standardise business processes (Sadowska, 2013),

- to document business activities, data and knowledge (Rockwell & Bajaj, 2005; Sadowska, 2013),

- to focus attention towards critical information elements (Fiol & Huff, 1992),

- to improve communication between stakeholders (Sadowska, 2013) leading to decreased miscommunication and conflicts,

- to facilitate brainstorming (Sadowska, 2013),

- to act as an external memory or repository framework and triggering prior knowledge (Fiol & Huff, 1992), and

- to support management initiatives (Nonaka, 2007; Sadowska, 2013).

These lower-level purposes can ultimately be aggregated into three high-level ones (De Meyer, 2015): communication (with a special ability to focus attention on critical parts), specification (to define a concept and show its relations to context elements), and referencing (the documentation and positioning of information into a repository framework in order to use for later communication).

However, we notice that all diagram purposes are human-oriented and leading back to communication and human problem solving as the final goal (Moody, 2009). Thus ultimately, the challenge is to optimise graphical representations so that they are optimally fit for their purpose of supporting communication.
Additionally, applying Garvin’s cascade approach (Garvin, 1984), we define qualitative graphical representation design as a cascade of three steps. Compliance to all three design steps is necessary to ensure the delivery of a qualitative diagram. First, the user preferences are determined (user-based approach), which are related to the communicative purpose of the representation. Secondly, these preferences are translated into product requirements/features (product-based approach). Lastly, the graphical representation is designed to meet the product requirements (production-based approach). The cascade approach emphasises that quality is not something that can be achieved by just applying a visual notation and spatial arrangement, but that it rather needs to be built (designed) into the diagram (Larkin & Simon, 1987).

1.3. The mole in graphical representation design: visual dialects

Today, a lot of the communication value of a graphical representation is still determined by the designer (Evitts, 2000). He has to deal with an abundance of available approaches towards quality, which results into a lack of directedness (Garvin, 1984; Rogers & Scaife, 1998) and disagreement on what makes a graphical representation ‘good’ (Moody, 1998; Rockwell & Bajaj, 2005).

Second, visual syntax has been systematically under-valued and put in the back seat. While evaluation techniques for visual semantics have already matured (Berenbach, 2004), a design rationale for visual notation and diagrams has been largely absent (Chen, 2005; Moody, 2009; Rogers & Scaife, 1998).

Lastly, the proposed guidelines for visual design that have been stated are mostly vague, complicated and opinion-based, lacking an underlying structure explaining how visual properties relate to one another (Nelson et al., 2012).

The lack of scientific basis for diagram design principles (Rockwell & Bajaj, 2005) is leading to engineers and other designers to pick and use visual languages and spatial arrangements based on habits and personal preferences (Moody, 2009; Sweller, Chandler, Tierney, & Cooper, 1990). This unthoughtful design rationale has led to the birth of countless visual dialects, which are often ineffective (Rogers & Scaife, 1998) as the effects of graphic design choices tend to be counterintuitive (Wheildon & Heard, 2005).

For example, CPN (Coloured Petri Nets) and PrTN (Predicate/Transition Nets) are only two of the many dialects of the Petri Net language (K Jensen, 1991; Kurt Jensen, 1980). The multitude of dialects has obstructed the exchange of net models within the field (Billington, Christensen, & Høe, 2003). Another example in the field of Requirements Engineering is i*, a
leading goal modeling language, which is absent of a design rationale for diagramming choices (Moody, Heymans, & Matulevicius, 2009). The vaguely defined syntax and semantics in goal modeling have led to diverse interpretations and multiple dialects, diminishing the usefulness of the languages (Woodman, 1988).

A second consequence of this lack of sound graphical design principles is that little progress has been made towards a holistic framework for improving or evaluating graphical representation design (Chen, 2005; Nelson et al., 2012; Rockwell & Bajaj, 2005; Scaife & Rogers, 1996). Instead of converging towards a scientific basis of visualisation techniques, the field seems to diverge into a multitude of visual dialects.

A first big contribution towards a scientific basis for quality in graphical representation design was made by Moody (2009) with his pioneer paper The Physics of Notations. His work provides nine principles for the design of visual notations, which has a major reference theory at the moment for graphical representation quality in Information Systems and Computer Science.

1.4. Scope and objective of this paper

1.4.1. Research question

In this paper, we aim to answer the question of what quality in graphical representation design entails. Our goal is to reveal the scientific basis of graphical representation design by formulating diagram design principles based on theory and empirical evidence. With this we hope to deliver a common ground and strong scientific foundation for visualisation techniques so that future research can be accelerated and the problematics of visual dialects subsides.

1.4.2. Research objectives

The goal is twofold. First, we aim to reveal the implicit knowledge behind existing rule-of-thumb guidelines and bring it together with the descriptive theoretical knowledge from reference domains. Second, we try to create sensitivity and awareness around the importance of proper design rationale and point out the scientific character of visual design.

As a long-term vision, we hope to kick-start a self-conscious design culture (Alexander, 1964) and catalyse future research in the field quality in graphical representation design.

1.4.3. Scope

This paper aims to build further on Moody’s work, but clearly differs from The Physics of Notations in that it provides principles for the use of visual notations. While the use of visual
notations for making diagrams is a sentence-level issue (bottom-left quadrant on Figure 1), the design of visual notations is a language-level issue (top-left quadrant of Figure 1). Both papers are thus complementary in nature and together lay the foundation for a scientific basis for quality in graphical representation design.

*Figure 1: While The Physics of Notation has focussed exclusively on the top-left hand quadrant, this paper will focus on the bottom-left quadrant of the figure. Source: Moody (2009)*
2. Methodology

2.1. Revealing the missing link: a holistic literature study of reference domains

To achieve our goal of revealing the implicit knowledge behind existing rule-of-thumb guidelines for diagram design and bringing it together with relevant theoretical knowledge, both prescriptive and descriptive literature was analysed. This enabled the identification of ‘the missing link’ (i.e., the missing principles) as exhaustively as possible.

Additionally, to structure the literature study, *Communication Science* was utilised as a guiding grid to cover all aspects of graphical representations as a communication tool.

**Prescriptive literature.** First, we aimed to reveal and gather the implicit scientific knowledge scattered and hidden in existing design guidelines by looking at prescriptive literature across the three aspects of the design-communication-decoding dimension. We expected guidelines in the field of documentation, teaching material and System Engineering to provide some useful insights.

**Descriptive literature.** Second, we gathered descriptive knowledge in order to scientifically evaluate and underpin the implicit knowledge discovered in the previous step. We expected cognitive theories to contribute heavily, because previous research had already confirmed that cognitive psychology is a promising foundation for the improvement of modelling languages (Rockwell & Bajaj, 2005; Rogers & Scaife, 1998; Zugal, 2013).

Using these two dimensions to structure the literature study (i.e., a descriptive-prescriptive dimension and an encoding-communication-decoding dimension) was assumed to ensure a holistic approach towards graphical representation design.

2.2. Formulating the missing link: the graphical representation design principles

After revealing the implicit knowledge and identifying their theoretical foundations in descriptive literature, the collected knowledge was integrated formally in the form of design principles. In order to do this, the collected prescriptive guidelines and descriptive theories were modularised into categories, after which overarching principles were extracted and formulated accordingly. The aim is to be exhaustive and as structurally coherent as possible in the formulation of the principles to minimise room for confusion or interpretation.
3. Design principles for graphical representations

3.1. A scientific basis for graphical representation design

As established above, the purpose of graphical representations is to facilitate the conveyance of information between stakeholders. Shannon and Weaver’s communication theory (Shannon & Weaver, 1949) argues that communication consists of five elements (see Figure 2): an information source, a transmitter who encodes the information, a channel which facilitates the communication of the message, a receiver which decodes the information and finally a destination (Fiske, 1990).

![Figure 2: Shannon & Weaver’s Communication Theory applied to graphical representations.](image)

The three phases of Shannon & Weaver’s Communication Theory, i.e. the encoding, communication and decoding phase, are used as matrix dimensions to structure the descriptive and prescriptive literature study (see Figure 3).

![Figure 3: Reference domain matrix](image)
3.1.1. Reference Domains

**Prescriptive literature.** Covering the prescriptive side, we investigated (1) for encoding practices, the domains of *Graphic Design* and *Aesthetics*, (2) for the information conveyance phase, the domains of *Requirements Engineering* and *Instruction Message Design*, and (3) for the decoding aspect, the domains of *Educational Sciences*, *Teaching Material Guidelines* and *Learning & Instruction Guidelines*.

**Descriptive literature.** Relevant descriptive knowledge was gathered in various fields related to graphical representation design in order to scientifically *evaluate* and *underpin* the collected guidelines. We covered (1) the encoding aspect with *Semiotics* and *System Engineering* research, (2) the communication aspect through theories from *Computer Sciences* and *Instructional & Educational Technology*, and (3) the decoding aspect by studying *Cognitive Science & Psychology* and *Brain Sciences*. Previous research confirms that concepts from *Cognitive Psychology* are a solid foundation for improving modelling languages and methods (Rogers & Scaife, 1998; Zugal, 2013).

Additionally, we searched for relevant literature in fields situated in between descriptive and prescriptive literature, such as *Visual Languages in Computing Science, Instructional Science & Educational Psychology* and *Text Processing Research*.

When situating these reference domains in the reference domain matrix (presented in Figure 3), it can be concluded that approximately all areas in the problem space of graphical representation design are thus covered.

3.1.2. Emerging graphical representation design principles

The descriptive and prescriptive knowledge from our literature study is synthesized and integrated in Table of Principles in Appendix A (first column from the left). In the second column, it is indicated for each theoretical model or prescriptive guideline which of *the 4 dimensions of graphical representation design* (see Section 1.2.1) it addresses. The gap between the descriptive knowledge and the prescriptive guidelines was bridged through the creation of a well-defined set of scientifically underpinned principles, which are presented in the third and fourth column of the table.

3.2. The 12 principles of graphical representation design

As mentioned, the literature study resulted in the identification of 12 principles for graphical representation design. Three *types* of principles have been distinguished: fundamental, practical and integrating principles. The *fundamental principles* provide the foundation for and the reasoning behind the seven *practical principles*, which in their turn discuss how to
achieve diagram effectiveness. Finally, the integrating principles indicate how to resolve conflicts between principles.

In what follows, the necessity and the usefulness of each of the 12 principles is illustrated with respectively theoretical knowledge and a discussion of related practical, yet research-based, guidelines for proper graphical representation design.

### 3.2.1. Fundamental Principles

Three fundamental principles were identified: the principle of diagram acquisition, the principle of minimality and the principle of complexity management.

#### Principle 1. Principle of focus on diagram acquisition

The principle of diagram acquisition states that a diagram engineer should focus on diagram acquisition and the cognitive processability of the diagram, whereas semantical completeness and correctness of the diagram are esteemed to be relatively less crucial.

In the past, quality frameworks for conceptual modelling have been characterised by a vigorous pursuit for semantical completeness and correctness. Two illustrations of this are leading quality frameworks of Lindland, Sindre and Sølvberg (LSS) (1994) and Bunge–Wand–Weber (BWW) (1990). However, immoderate focus on completeness and correctness often translates into complex diagrams which trigger split-attention effects (Chandler & Sweller, 1991, 1992; Van Merriënboer & Kester, 2005) and increase cognitive load (Sweller, 1988). These diagrams have low readability and thus fail to achieve their purpose: facilitating and enhancing conveyance of information (Cierniak, Scheiter, & Gerjets, 2009; Florax & Ploetzner, 2010; Ginns, 2006; Rockwell & Bajaj, 2005; Sweller et al., 1998; Zugal, 2013). Notice that when this happens, users also benefit less from engineering efforts for optimising semantical completeness and correctness. Consequently, in order for graphical representations to fulfil their raison d’être, diagram engineers should primarily focus on diagram acquisition and cognitive processing, whereas semantical completeness and correctness are relatively less crucial. The question in then how to design for optimal diagram acquisition.

**Descriptive theories.** Research in *Cognitive Science* has found that on average, the capacity of working memory is limited to processing 7 plus or minus 2 elements at a time (Miller, 1956). According to more recent work, only 3 to 5 elements can be simultaneously handled for information processing (Cowan, 2010). When this working memory capacity is exceeded, learning is affected and cognitive processing becomes ineffective (Sweller, 1994).
Graphical representations should thus be designed to accommodate this limited human cognitive processing capacity.

Cognitive Load Theory (CLT) states that the cognitive load of a graphical representation is caused by both the intrinsic nature and complexity of the material (the intrinsic cognitive load), and the way the material is presented (extraneous cognitive load) (Sweller et al., 1998). While intrinsic cognitive load cannot be changed by graphical design as it is intrinsic to the represented data, extraneous cognitive load can be minimised by improving the instructional design (Sweller et al., 1998). In domain literature, the ability of graphical representation design to reduce the cognitive load of an information-equivalent representation, is called computational offloading (Larkin & Simon, 1987).

**Guideline: Adhere to the principles of minimality and complexity management.** To maximally benefit from computational off-loading opportunities, we argue graphical representation design should adhere to two additional fundamental design principles: the principle of minimality and the principle of complexity management, which are discussed hereafter. Both principles address the issue of diagram overcrowding and provide more practical guidelines for their specific focus.

**Principle 2. Principle of minimality**

According to the principle of minimality, graphical representations should contain the minimum possible amount of symbol types.

Due to the growing complexity of data (i.e. an increasing level of abstraction and interconnectedness of data), which is aggravated by the growing volume, variety, velocity and veracity of data, visual representations often get overcrowded (Ellis & Dix, 2007). Too much information needs to be fit on a physically and cognitively constrained display space. The issue of diagram overcrowding is two-sided: on the one hand, it is caused by an excessive amount of types of graphical symbols on the diagram, on the other hand, by an exaggerated number of symbol instances. Visualisation techniques need to adopt strategies for dealing with both sides of diagram overcrowding. In the principle of minimality, the issue of excessive amounts of symbol types is addressed, while the issue of the excessive use of instances is discussed under the principle of complexity management later on (see Principle 3 below).

The first way to avoid visual overcrowding and thus ensure diagram readability, is to adhere to the notion of diagram minimality. Minimality in element types, e.g., non-redundancy of
activity types on BPMN models, increases diagram readability (Si-said Cherfi, Akoka, & Comyn-Wattiau, 2002). In what follows, practical guidelines collected from various fields are presented to achieve diagram minimality.

**Guideline 1: Only model the strictly necessary (semantics).** Prescriptive guidelines in *Software Engineering* prompt to respect *diagram determinism* (Zugal, 2013) by only modelling the strictly necessary semantical constructs in order to ensure message conveyance (Ambler, 2005; Sadowska, 2013) and to enable proper modularisation (see also Principle 5 below) (Zugal, 2013).

**Guideline 2: Resist the temptation to add embellishments (syntax).** *Learning & Instruction* guidelines emphasise to resist the temptation to add variables and symbol types extraneous to learning to supposedly increase clarity (Schnotz, 2013). Brophy & Good (1986) argue that the higher the number of redundant feature and element types on a diagram, the more difficult learning becomes. Excluding these redundant symbol types, i.e., adhering to the *guideline of coherence* (Van Merriënboer & Kester, 2005), has found to positively impact information transfer and *elaborative learning* (Van Merriënboer & Kester, 2005).

**Illustration of diagram minimality in practice: Visual filtering.** In practice, reduction of clutter in the field of *Data Science* is achieved by techniques such as *visual filtering*, which filters dataset query results and suppresses unrequested (i.e. redundant) information (Ellis & Dix, 2007; Koshman, 2010).

**Principle 3. Principle of complexity management**

According to the *principle of complexity management*, diagram complexity should be managed by limiting the number of element instances on the diagram.

The second and complementary way to avoid visual overcrowding is to adhere to the principle of complexity management. Where the principle of minimality is concerned with the number of different types of graphical symbols on a graphical representation, *complexity management* is concerned with the number of graphical instances used in one diagram.

**Underlying descriptive theories.** Although the very purpose of graphical representations is to facilitate communication and learning by computational off-loading (Van Merriënboer & Sweller, 2005), they themselves are often still too complex (Tversky, B., Morrison, J. B., & Betrancourt, 2002) and get overcrowded with a large amount of symbol instances. *Learning & Instruction* literature confirms that by utilising an excessive number of instances in one
diagram, intrinsic cognitive load gets too big and complexity management measures are needed to reduce the demands on working memory (ref. cognitive load theory). Literature in Psychology and in Educational Science provide additional, motivation-related reasons for complexity management. First, recent research in the field of Educational Psychology has found that learning material, which is perceived to be more supportive, is able to improve learning by reducing fear of failure and increasing their self-efficacy (i.e., the confidence in one’s own ability to achieve intended results) (Cennamo, 2016). Additionally, higher levels of motivation lead to improved self-monitoring and deeper learning strategies (Pintrich, P. R. & Schunk, 2001; Sweller et al., 1998). In short, the general message for designers is that less can be more (Schnotz, 2013). In what follows, practical guidelines for managing diagram complexity are provided. Additionally, an illustration of complexity management in the field of Information Systems & Computer Science is given.

**Guideline 1: Prevent overlap and instance redundancy.** Guidelines from Software Engineering and Computer Science advice to prevent overlap and instance redundancy (Ambler, 2005).

**Guideline 2: Restrict the number of instance(s) (groups) to 7 plus or minus 2.** The ensemble of instances on the diagram should be kept manageable to facilitate schema acquisition (Moreno, 2006). Following Miller (1956), maximally 5 to 9 instances should be put on a diagram to be simultaneously processed and integrated.

**Guideline 3: Limit diagrams to one page.** Finally, complexity management is not only desirable from an effectiveness and efficiency point of view. Research in Computer Science has found that often people prefer single-page diagrams (Ambler, 2005) and simplified information delivery (Gaissmaier et al., 2012), which is related to a motivational aspect of the diagram readers.

**Guideline 4: Integrate the different sources of information.** Lastly, multimedia-oriented research in Instructional Science has found that integrating multiple sources of information (e.g., a multitude of diagram instances) which are self-explanatory or redundant into one non-redundant diagram, promotes information transfer and learning (Van Merriënboer & Kester, 2005).

**Illustration of complexity management in practice.** In Information Systems & Computer Science, BPMN model engineers often design diagrams according to the 7 Process Modelling Guidelines (7PMG) (Mendling, Reijers, & van der Aalst, 2010) and the guidelines of Bruce Silver (Silver & Richard, 2009). The resulting diagrams contain as few element
instances as possible, have only one start and end event, fit onto one page (however, models with more than 50 elements are decomposed).

3.2.2. Practical Principles

So far, three fundamental principles of graphical representation design have been established. Now, eight practical principles and corresponding practical guidelines are presented to optimise graphical representation design for diagram acquisition. Three types of practical principles are distinguished: the grouping principles (Principle 4 and Principle 5), ordering principles (Principle 6 and Principle 7) and visualising principles (Principle 8, Principle 9 and Principle 10).

Principle 4. Grouping - Principle of chunking

The principle of chunking states that information on a graphical representation should be chunked into meaningful parts in order to improve understanding and diagram acquisition.

Underlying theories. As explained above, research in Psychological Science has found that the working memory is limited and can only successfully process three to five meaningful items at a time (Cowan, 2010; Rowlatt, 2008). Literature in the field of Cognitive Psychology describes an experiment by Bower (1970) who found that learning is improved when material is organized into categories, facilitating learning through the category interrelationships. This is in line with an experiment in the field of Learning & Instruction. Here, Lowe (2003) finds that clear visual-spatial characteristics (i.e., structural coherence) on diagrams positively affect information extraction, while representations that lack structural coherence cannot support information extraction. These experiments prove the effectiveness of visually segmenting material into meaningful groups to improve learning from graphical representations (Van Merriënboer & Kester, 2005; Westelinck et al., 2005).

The method of meaningfully splitting and compressing the information in separate units so that it can fit the available working memory capacity, is called chunking. Chunking is a form of non-hierarchal grouping, also referred to in Information Systems and Computer Science as ‘horizontal modularisation’ (De Meyer, 2015). The technique of chunking fastens feature recognition, value lookup and inference making (Larkin & Simon, 1987). Additionally, chunking allows to make diagrams structurally coherent in an highly cognitively effective way through avoiding the use of labels to denote connectivity between elements (Larkin & Simon, 1987).
Gestalt theory (Wertheimer, 1923) supports the use of visual techniques, and thus chunking, to indicate item relatedness in a graphical representation (Koshman, 2010). The theory argues that the perception of objects in an environment is not defined by the sum of the individual instances, but rather by the total configuration of the elements together as a whole.

In what follows, practical guidelines for and illustrations of chunking are provided.

**Guideline 1: Group diagram information into three to five units.** Practically, chunking can be done by first distinguishing advisably three to five categories of information. Although research indicates that the optimal number of chunks is around three to five (in correspondence to the working memory capacity), the amount of information per unit is not particularly constrained in size and complexity (Sweller, 1988).

**Guideline 2: Express category distinction through the use of one or multiple Gestalt laws.** After the identification of different chunks of information, visual variables and techniques can be used to express category distinction. Following Gestalt theory, we distinguish four ways to chunk information in a display area: by proximity, similarity, symmetry or common fate (Koffka, 2013).

(1) **Chunk by spatial proximity**

The law of proximity in Gestalt Theory states that the human mind perceives objects that are close together as a group of related items. Figure 4 illustrates this the law of proximity: the figure shows 72 circles, yet we automatically perceive the left side of the figure as a group of circles and the right hand side as another groups of circles.

![Figure 4: Illustration of the Gestalt Law of Proximity](image)

Chunks by spatial proximity is often recommended in Graphical Design guides (Bertin, 1981), and is frequently applied in the field Information Systems and Computer Science (Colin Ware, 2005).

(2) **Chunk by similarity**

To chunk by similarity, the designer can make use of visual variables like colour, shape and
texture to indicate element relatedness without physically connecting them with arrows (Figure 4: Illustration of the Gestalt Law of Proximity) (Koshman, 2010).

![Figure 5: Illustration of the Gestalt Law of similarity](image)

In Data Science, colour is often used to highlight datasets which share mutual attributes to facilitate correct data selection (Koshman, 2010; C. Ware, 2004).

(3) Chunking through symmetry

Chunking is also possible by positioning objects symmetrically around a focal point. Gestalt Theory states that the human mind connects symmetrical elements even though they are not physically connected (Figure 6: Illustration of the Gestalt Law of Symmetry: individuals are more likely to observe three pairs of brackets rather than six individual ones). If the objects are similar, the chances are higher they will be recognised as related symmetrical elements.

![Figure 6: Illustration of the Gestalt Law of Symmetry](image)

(4) Chunking by common fate

Objects that are on the same ‘trend line’ or have the same orientation or direction tend to be perceived as a group as whole. By giving elements the same orientation, their similarity can be communicated.

**Guideline 3: Use visual variables to enhance category distinction.** Visual variables like colour and shape can be used to further enhance category distinction. Ware (2005) promotes the use of a combination of contour, colour, motion or texture to segment the display space into regions. For an overview of practical guidelines for the use of visual variables to increase diagram expressiveness, we refer to the Visualising - Principle of expressiveness (principle 10).

**Illustration of the use of chunking in practice.** In Information Systems and Computer Science, designers often use container elements, i.e., a combination of spatial proximity and
visual variables, to spatially group information together. An illustration hereof is the use of lanes in Business Process Models to distinguish different functional departments.

**Principle 5. Grouping - Principle of modularisation**

| The principle of modularisation states that the intrinsic cognitive load of graphical representations can be decreased by using hierarchical structures to represent element relationships. Either a top-down (decompositioning) or bottom-up (summarisation) approach can be taken. |

**Underlying theories.** Especially in the field of *Information Systems & Computer Science*, which is characterised by high complexity levels (Chen, 2005; Dijkstra, 1972; Ganek & Corbi, 2003), the cognitive load of diagrams can be excessively large and exceed the capacity of working memory. At first instance, a diagram engineer tries to reduce a diagram’s extraneous cognitive load through design improvement. However, often extraneous cognitive load cannot be decreased any further and the only option is to reduce the intrinsic cognitive load (i.e. intrinsic complexity) of the information itself. To do this, the technique of modularisation is proposed by both research in *Instructional Science* (Gerjets, Scheiter, & Catrambone, 2004) and *Information Systems & Computer Science* (Ambler, 2005; Regnell et al., 1996). The technique of modularisation entails first chunking the information into subcomponents (see the Grouping - Principle of chunking above) and then add hierarchy to the diagram by making relations between diagram elements explicit.

Some benefits are named for modularisation in literature. First, clear visual-spatial characteristics (i.e. structural coherence) have been found to help see relationships between elements (Hall & O’Donnell, 1996; Krajcik, 1991; Linn & Muilenburg, 1996) and support learning from graphical representations (Lowe, 2003; Paas et al., 2004; Saltz, 1971; Westelinck et al., 2005). This is especially the case for information of high cognitive load (Richard Cox, 1999; Haygood & Bourne, 1965; Mayer, 1976; Neisser & Weene, 1962) and when the user is a novice (Sweller, 1994).

Second, research in Cognitive Psychology has found that creative thinking with diagrams requires an hierarchical structure and an easy identification of cross-links (Novak & Cañas, 2008), which modularisation can provide.

---

3 In Information Systems and Computer Science, the term ‘modularisation’ is sometimes used to refer to the technique of nesting, where control elements like gates are used to split and gather different pathways in a diagram. We, however, define modularisation as the hierarchical grouping of information, also referred to as ‘vertical modularisation’ in Information Systems and Computer Science (De Meyer, 2015).
Next, research in Educational Psychology by Mayer & Mautone (2007) found that users who received structural graphical representations were able to formulate more relational (however not necessarily more causal) statements than the users in the control group.

Furthermore, hierarchical diagrams allow for simultaneous learning. In the case of high-element interactivity, simultaneous display is often required to support simultaneous processing and manipulation (Bourne, L. E., Ekstrand, & Dominowski, 1971; Ginns, 2006; Paas et al., 2004).

Lastly, modularisation positively impacts the maintainability of the diagram. For example, in Information Systems and Computer Science, software is modularised in packages to improve maintainability (Bavota, De Lucia, Marcus, & Oliveto, 2010).

In short, the positive impact of modularisation seems to be very stable over a range of different conditions like different measures of learning time, low versus high prior knowledge, varying learning tasks and the extent of additional explanations (Gerjets et al., 2004). However, in order to be able to modularise, engineers need a visual notation that offers the necessary semantical constructs to do so (Ågerfalk et al., 2007; Moody, 2009). Examples of modularisation constructs are ‘subsystem constructs’ (like subprocesses seen in BPM models) (De Meyer, 2015; Sadowska, 2013) and ‘decomposable constructs’ (like packages in Object Oriented software which group related classes together) (Bavota et al., 2010). Although literature in various fields agrees on the desirability of modularisation (Bavota et al., 2010; De Meyer, 2015; Mautone & Mayer, 2007; Saltz, 1971), many leading visual notations still lack the required modularisation constructs. An example is i*, a leading visual notation in goal modeling, which does not allow for proper modularisation as it does not offer semantical constructs.

Still, the presence of such semantical constructs is not enough: also on a syntactical level, clear diagramming principles and solutions for modularising must be defined (Moody, 2009). In what follows, we present initial modularisation guidelines for both a top-down and a bottom-up approach (Regnell et al., 1996).

*Table 1: Modularisation Principle: Overview.*
### A. Guidelines for top-down modularisation

The first way to modularise is to take a top-down approach, where the most inclusive and often abstract concepts are decomposed into their lower-level, often more specific, subcomponents (Novak & Cañas, 2008). Decomposing can be diagrammed by using outgoing arrows from the higher to the lower levels, and optionally, using control constructs (like split-gates in BPM).

<table>
<thead>
<tr>
<th>Modularisation principle</th>
<th>Convention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top-down approach</strong></td>
<td></td>
</tr>
<tr>
<td>Decomposition</td>
<td>Adhere to explicit complexity limit of 7 +/- 2 bubbles</td>
</tr>
<tr>
<td>Graduality</td>
<td>Perform a fan-in/fan-out balance check</td>
</tr>
<tr>
<td>Fluency</td>
<td>Balance control constructs with diagram fluency</td>
</tr>
<tr>
<td><strong>Bottom-up approach</strong></td>
<td></td>
</tr>
<tr>
<td>Abstraction</td>
<td>Summarise by backward and/or forward fading</td>
</tr>
<tr>
<td>Recursive decomposition</td>
<td>Create black-boxes that allow for reuse</td>
</tr>
<tr>
<td>Strong cohesion</td>
<td>Balance the ratio of abstraction &amp; fragmentation</td>
</tr>
</tbody>
</table>

**Guideline A.1.: Gradually increase element refinement.** While decomposing the diagram, it is important to gradually increase element refinement so that the element relationships are obvious (Van Merriënboer & Sweller, 2005). A technique to measure this graduality, is to count the fan-in and fan-out of every module (component), where fan-in represents the amount of other modules ‘calling’ the specific module and fan-out the number of other modules that are called by the designated module (Gruhn & Laue, 2006). Simply put, the technique compares the number of arrows arriving and the number of arrows leaving a module. Modules with a large fan-in are usually small, low-level submodules, while modules with large fan-out are mostly larger modules at the top of the hierarchy levels (Gruhn & Laue, 2006). When a module with both high fan-in and high fan-out is encountered, this signals that the diagram should be further improved (Gruhn & Laue, 2006).

**Guideline A.2.: Keep the depth of decomposition manageable.** As for nesting depth, it is important to adhere to the explicit complexity limit of 7 +/- 2 units per diagram in correspondence to the limitations of human short term memory (Miller, 1956). Furthermore, respecting the explicit complexity limit is also important to maintain processing fluency (fluency of thoughts) (Kuhn & Stahl, 2003; Rolf Reber, Schwarz, & Winkielman, 2004) and motivation of the reader (Cennamo, 2016; Pintrich, P. R. & Schunk, 2001; R. Reber, Winkielman, & Schwarz, 1998).
B. Guidelines for bottom-up modularisation

The second approach to modularise information is to start from the concrete concepts and work your way up to the abstract one by the process of summarisation (Berenbach, 2004; Moody, 2009). Summarisation allows to *fade* (hide or omit) certain chunks of information and provide an overview without cluttering the diagram with details (Regnell et al., 1996). The modularised diagram structure is foldable, permitting to show the diagram at different levels of refinement (Berenbach, 2004). Fading can be done both upstream or downstream: forward fading will hide a ‘mother element’, while backward fading will omit a ‘daughter element’ (Paas et al., 2004). In the case of tree diagrams, fading can also be done at a branch-level (i.e. omitting intermediate branches) or at an abstraction-level (omitting complete intermediate levels) (Figure 7: Illustration: fading at branch level. Source: Ding & Mateti (1990)Figure 8) (Ding & Mateti, 1990).

![Figure 7: Illustration: fading at branch level. Source: Ding & Mateti (1990)](image)

![Figure 8: Illustration: fading at abstraction-level. Source: Ding & Mateti (1990)](image)

In literature, the strategy of fading has been found to improve understanding and decrease unproductive searching and learning activities (Paas et al., 2004).

**Guideline B.1.: Introduce black boxes to increase diagram complexity.** Through *recursive decomposition* (DeMarco, 2002), smaller schema’s are fitted into larger schema’s, allowing to turn lower level information into black-boxes. The use of black-boxes benefits communication because it guides attention towards the critical parts of the representation.

---

4 We use the term ‘recursive decompositioning’ as a synonym for the process of summarisation.
and frees up working memory capacity (Paas et al., 2004). Additionally, it avoids duplication of work by allowing reuse of schema parts (Gruhn & Laue, 2006; Regnell et al., 1996), which according to schema theory, increases user performance (Paas et al., 2004).

**Guideline B.2.: Ensure strong diagram cohesion.** Last, it is important to ensure strong diagram cohesion by carefully balancing the level of abstraction and fragmentation. A coherent diagram ideally keeps related elements together but fosters abstract thinking (Zugal, 2013).

**Illustration of modularisation in practice.** In *Software Engineering*, modularisation is used to facilitate and improve software design, coding and software maintainability.

As illustrated, the principle of modularisation has many benefits but still lacks some theoretical backbone on both semantical and syntactical level. With our diagramming conventions for top-down and bottom-up approaches, we hope to have provided an overview and first basis for syntactically effective modularisation solutions on which future research can build further.

**Principle 6. Ordering - Principle of direction of information**

| The Principle of direction of information states that the readability of a graphical representation is improved by maintaining continuous and consistent control flow direction. |

**Underlying theories.** Continuity is an important factor in the understanding of graphical representations (Colin Ware, Purchase, Colpoys, & McGill, 2002). The Gestalt *law of good continuation* describes the human ability to integrate or connect separated components on a graphical representation under the condition that the flow direction is clear and continuous (Field, Hayes, & Hess, 1993). An illustration of the law of good continuation can be found in Appendix B (Figure 19).

**Guideline 1: Align graph elements on continuous and smooth paths.** In order for diagrams to benefit from this human cognitive ability of continuation, related graphical elements should be aligned along a common diagram path which is preferably continuous and smooth (e.g. no zigzagging or brisk, unexpected turns) (Figure 9) (Field et al., 1993).
Guideline 2: Draw edges along orthogonal vertices. Following research in *Information Systems & Computer Science*, straight lines, either vertically or horizontally, are easier for a user to follow (Ambler, 2005). Diagram elements should be modelled along horizontal and vertical lines, as if they were put on the unit grid points of the diagram (Ambler, 2005; Irani & Ware, 2003). According to Tamassia et al. (1988), for drawing hierarchical diagrams, it is advantageous to (temporarily) place parallel vertices on the design space so that edges can be consistently drawn along those lines, from top to bottom, in the same direction.

Guideline 3: Maintain consistent control flow direction. Prescriptive literature in *Information Systems & Computer Science* (Ambler, 2005; Becker, Rosemann, & Uthmann, 2000; Long, 2002) recommends to maintain consistent control flow direction within and across diagrams. As in the Western world the general reading direction is from left to right, direction of information should ideally be from left to right and additionally, from top to bottom. The importance this principle is illustrated by Figure 10. Lastly, the guideline of consistency in flow direction also applies for the direction of arrows, text labels and shapes of container elements (in case the latter indicate direction) (Evitts, 2000).

Illustration of the application of consistency in information flow in practice: the N-squared chart. This graphical design principle of direction of information is already well-applied in various fields. An example in the field of System Engineering is the N-squared chart.
chart, where input data enters on the top-left side of the diagram, while output leaves at the bottom-right corner (Long, 2002). In the field of Project Management, baseline schedules take on a left-to-right, top-to-bottom approach (Vanhoucke, 2012).

**Principle 7. Ordering - Principle of internal and external linkage**

The principle of internal and external linkage states that learning from graphical representations can be improved through internal linking of diagram elements and through linking the diagram to the external context.

**Underlying theories.** Mental integration of information is a critical antecedent of learning, but puts high demands on working memory (Chandler & Sweller, 1992). Research in *Experimental Psychology* finds that the better integrated learning material is, the greater long-term memory and resistance against interference (i.e., other material interfering with the memorised information leading to decreased recall and correctness of recall) (Houston, 1965; Saltz, 1971; Sweller, 1988). The integration of information is twofold: the diagram information should be internally linked (connecting subcomponents within the material), but preferably also linked to relevant external elements (connecting to the external context). When material is separated over either display space, time or multiple information sources, learners must divide their cognitive capacity over the different information sources (Ginns, 2006), and a so-called split-attention effect can arise. However, empirical testing found that the split-attention effect only applies for low-interactivity information, not for complex information (Cierniak et al., 2009). Therefore, presenting information in a fully integrated format greatly benefits learning of low-interactivity information, but not necessarily learning of complex information. Therefore, separate practical guidelines with illustrations are presented for diagramming information of varying complexity.

**A. Internal linkage**

**Guideline A.1.: Model the complete diagram breadth first, drill down afterwards.** In the first phases of graphical representation design, drilling down too soon into low-level details of the model, might lead to losing track of the primary goal and endangers the overarching logic and structure of the graphical representation (Berenbach, 2004). Therefore, the first graphical design effort should be holistic and focussed at the complete diagram breadth (Berenbach, 2004). Also, scoping the diagram on the entire knowledge domain and deciding what the diagram will and will not cover, will help not to lose track of the design goal and diagram message.
Guideline A.2.: For graphically representing low-complexity information, make relationships between diagram elements explicit. For low-complexity information, empirical evidence encourages to make the relationships between elements explicit, arguing that the scattering of elements on a diagram leads to decreased structural coherence (Caillies, Denhière, & Kintsch, 2002; Rockwell & Bajaj, 2005). By decreasing the need for mental integration by providing a well internally-integrated graphical representation, working memory capacity can be freed from integrative processing aspects and instead used for task performance (Carlson, Khoo, & Elliott II, 1990; Ginns, 2006). Similarly, research in *Experimental Psychology* recommends to integrate multiple mutually referring information sources (e.g., text with complementary graphical representations) as much as possible, instead of fragmenting them for aesthetical reasons (Sweller et al., 1990).

Guideline A.3.: For graphically representing complex information, make use of text segmentation and labelling instead of spatial integration or linking. For high-complexity information, evidence in *Information Systems & Computer Science* states that weak coupling, i.e., minimizing the number of connections between instances, can be beneficial (Zugal, 2013). Less interactivity between modules means less potential (distracting) switches between modules (Zugal, 2013). The absence of spatial integration can be largely accommodated by segmenting text and labelling pictures (Cierniak et al., 2009; Florax & Ploetzner, 2010; Ginns, 2006). An illustration can be found in Appendix D.

B. External linkage

Linking the internal information of a diagram to relevant external environment (external linkage) benefits the understandability of the diagram (Berenbach, 2004). External linkage can be achieved by activating prior knowledge, contextualisation, and making use of repleteness.

Guideline B.1.: Activate prior knowledge. Research in *Cognitive Psychology* has found that, even though users might have prior knowledge which can help them understand new material (Gemino & Wand, 2003), they are often not able to recognize it themselves (Paris & Lindauer, 1976; Spires & Donley, 1998). Prior knowledge activation is therefore recommended in order to retrieve relevant existing knowledge from long-term memory and place it in working memory (Ormrod, 1999). This can be done through providing a cognitive map (Figure 11) of specific prior domain knowledge.
Guideline B.2.: Provide an overview. Providing an overview of, i.e., contextualising, how new material relates to and scopes onto the user’s existing knowledge and others problem domains, has found to positively impact learning and information transfer (Ormrod, 1999; Spence, 2007; Sweller et al., 1998; Van Merriënboer & Kester, 2005). Additionally, in some cases, our concept knowledge is based and dependent upon context (e.g., a cup might become a vase when there’s flowers in it) or a variety of examples, rather than on one single isolated exemplar (Ormrod, 1999). Context diagrams (Figure 12) or cognitive maps (Figure 11) are effective ways to provide users with an understanding of the different frames of reference, trigger memory and act as a supportive external memory (Fiol & Huff, 1992; Siau & Tan, 2005). They can also facilitate navigating between different graphical representations by making the link between the different sources of information clear.
**Guideline B.3.: Make use of repleteness: use familiar modelling languages.** In *System Information & Computer Science*, engineers explicitly make use of repleteness (i.e., familiarity and existing knowledge) in their models to increase diagram quality (Westelinck et al., 2005). It is recommended that engineers use modeling languages which are similar to what users already know or with which they are familiar, to facilitate understanding (Westelinck et al., 2005).

**Guideline B.4.: Make use of analogically reusable patterns.** Various authors promote analogical reuse of patterns in conceptual modeling design to ease diagram acquisition (Maiden & Sutcliffe, 1992; Nelson et al., 2012; Snoeck & Poels, 2000). The benefit in analogical reuse is that every time a pattern is applied successfully, an accumulation of learning processes lead to internal strengthening of the pattern (Carlson et al., 1990; Van Merriënboer & Kester, 2005). By making use of familiar patterns, the engineer allows for automation to take place, which improves information acquisition or task performance (Carlson et al., 1990; Van Merriënboer & Kester, 2005). An illustration of such a reusable pattern is the typical resource-event-agent pattern, which has been proven to be effective for UML and is often used in goal modeling (Maes & Poels, 2007).

**Guideline B.5: Use extremes of well-known dimensional sequences in word choice, visual variables and examples.** Additionally, research in *Cognitive Psychology* has found that the mind seems to store incoming information in the form of previously mastered cognitive structures (Saltz, 1971). Learning information elements that are situated at end points of a dimension (e.g. ice-cold versus hot) was found to be the easiest, while information coming from the middle of the dimension was found to be the most difficult (Pollio, 1968). Therefore, it might be beneficial to use extremes of well-known dimensional
sequences or cognitive structures in the use of examples, word choice and visual variables. An example of this would be to use the colour red to give a ‘bad performance’ or ‘alarm’ annotation to a symbol instance, while using green to indicate well-performing objects.

**Guideline B.6.: Use stretch text.** Stretch text (i.e., explanatory text) can be a way to provide some contextual or additional background info for novices. It can be added as a text label, or it can be hidden and shown as a ‘fold-open’ information icon. Ambler (2005) describes the use of *summary notes* (or so called *legends*) in UML diagrams, which are used to describe the purpose and the broader context of diagrams (Figure 13).

![Figure 13: Illustration of a summary note of a UML diagram. Source: Amber (2005)](image)

The advantage of stretch text is increased control over presented information and an increasingly readable diagram (Boyle & Encarnacion, 1998). Warmer & Kleppe (2003) describe the use of OCL (Object Constraint Language) to add constraints and stretch text to conceptual models while keeping cognitive diagram complexity manageable (Figure 14).

![Figure 14: Illustration of stretch text: the use of OCL on a UML diagram (indicated in red). Source: Ambler (2005)](image)

**Principle 8. Visualising - Principle of discriminability**

The **principle of discriminability** states that graphical representation readability can be improved by making diagram elements more perceptually discriminable.
Underlying theory. Building perceptual discriminability into a graphical representation is important, as perceptual processing is the first step in the processing of a graphical representation, before the cognitive processing can start.

![Diagram of perceptual and cognitive processing](image)

*Figure 15: Human processing of graphical representations. Source: Moody (2009)*

There are various ways to improve the perceptual discriminability of graphical representations. In what follows, we will indicate nine techniques for increasing the perceptual discriminability of a diagram.

**Guideline 1: Filter the represented content.** To start, it is important not to overload graphical representations with elements: the more instances, the more difficult to properly distinguish between them distinguish core from details and the higher the extraneous load (Paas et al., 2004) (see Principle of complexity management above). One way to avoid overcrowding a diagram is to filter the content you want to visualise (visual filtering). In the case were it is possible to omit or hide certain pieces of information, a distinction between core and details can be made, which will increase diagram effectiveness as users will be able to directly distinguish the critical from the non-critical (supportive) information elements (Ellis & Dix, 2007; Koshman, 2010).

**Guideline 2: Consider using textual differentiation.** In the case were completeness is judged to be key, textual differentiation can be a solution: omitting certain information from the diagram and putting it instead in a complementary text. However, this strategy of textual differentiation should only be used for high-complexity information, as for low-complexity information split-attention effects will occur (Cierniak et al., 2009).

**Guideline 3: Use modularisation.** The use of decomposable modular structures (see Grouping - Principle of modularisation above) allow to first only show the high-level elements (high perceptual discriminability), and afterwards increase the level of refinement (lower perceptual discriminability), which improves the perceptually discriminability of the graphical representation.

**Guideline 4: Ensure sufficient spatial distance between diagram elements.** Spatiality is a decisive factor in the level of discriminability of a graphical representation (Buagaja, 2001): distance between diagram elements ensures that the mind perceives them as separate identities (Ding & Mateti, 1990a). Therefore, literature in *Information Systems &
Computer Science prompts to avoid many close lines, spatial overlap of shapes and the crossing of lines (Ambler, 2005; Purchase, 2002; Colin Ware et al., 2002). If crossing lines is unavoidable, use 'jumps' (Figure 16) to make the lines more distinct and improve readability (Ambler, 2005).

![Figure 16: Illustration of depicting a crossing of lines as a jump. Source: Ambler (2005)](image)

Guideline 5: Maximise edge and node orthogonality. Additionally, literature in Information Systems & Computer Science argues to maintain a minimum angle between edges (i.e. lines, arrows) leaving a node to maximising edge orthogonality (Figure 17), and to fix and spread nodes on an imaginary ‘orthogonal unit grid’ to maximise node orthogonality (Purchase, 2002; Schuette & Rotthowe, 1998; Colin Ware et al., 2002).

![Figure 17: Illustration of the minimum angle principle: the edges on the right hand side (90°) are better discriminable than the left hand side (acute angle). Source: Ware et al. (2002)](image)

Guideline 6: Spread nodes uniformly of the display space. Lastly, Schuette and Rotthowe (1998) argue that uniform density of nodes improves diagram quality, i.e., the location/placement of instances should be spread over the full display area.

Guideline 7: Use barriers to separate object groups. It is beneficial to distinguish groups which share common features to increase perceptibility of the diagram (Van Merriënboer & Kester, 2005). The principle of chunking and spatial proximity (or distance) can be used to increase discriminability between different diagram parts (Paas et al., 2004). Practically, following the laws of Gestalt Theory (Koffka, 2013), increasing perceptual discriminability of separate chunks can be done by using barrier elements like to spatially separate them like lines or container elements (Ding & Mateti, 1990; Koshman, 2010).

Guideline 8: Avoid diagonal or bended lines. Diagonal or curved lines lead to lower diagram discriminability and should therefore be avoided (Ambler, 2005; Ware et al., 2002).
Guideline 9: Use reasonable and consistent element sizing. Make sure that the depiction size is reasonable for proper cognitive perception and reading with moderate effort. Research in *Metadata Applications* has yielded new techniques to increase visual discriminability of graphical representations, like the *fish-eye technique*, which enlarges icons as they are browsed (Koshman, 2010). Additionally, Ambler (2005) argues to keep consistent sizing of graphical symbols in order to not draw unwanted attention to larger instances.

Guideline 10: Use easily differentiable shapes and use them consistently. Research in *Cognitive and Behavioural Science* has found that graphical representations with elements having less differentiable shapes (i.e. shapes with lower perceptual differences) took longer and are harder to process (Peebles & Cheng, 2003). Similarly, Moody (2009) argues that shape is the most important way for a user to distinguish between different semantical constructs. Therefore, a diagram engineer should use a visual notation with perceptually easily distinguishable symbol shapes to increase readability of the diagram (Ding & Mateti, 1990).

Guideline 11: Use labels & diagram legends. Make use of *labels* to distinguish between symbol instances (but not symbol types, however) and to explain their meaning (Moody, 2009). To distinguish between symbol types, to accommodate users who are not familiar with the visual notation, offer a *diagram legend* which discusses all symbols used (visual syntax) together with their meaning (semantics) (Ambler, 2005).

Guideline 12: Use colours to increase element distinctiveness. In *Data Science*, distinctive item groups in datasets are made apparent through the use of colours (Koshman, 2010; Ware, 2004). If a visual language has low discriminability in symbol shapes, colour and other visual variables can be used to make the difference between diverse object types more apparent (Moody, 2009). In this case, add a legend to discuss the meaning of each colour used.

**Principle 9. Visualising - Principle of clarity**

The Principle of clarity states that diagram readability can be improved by ensuring clarity in semantics, syntax, naming and diagram structure.

**Underlying theories.** Experiments in the field of *Learning and Instruction* have found that learners often struggle with understanding graphical representations because they have no knowledge of the visual notation or general syntax that is used (Cox, 1999; Westelinck et al.,
Moreover, it was found that complex reasoning with unfamiliar visual syntax is prone to be ineffective (Sweller et al., 1998) and that users learn best from diagrams with clear visual properties (Lowe, 2003). Furthermore, research in Computer Science confirms that reasoning with diagrams like ERs requires an thorough semantical and syntactical understanding of the constructs and visual properties on that diagram (Cox, 1996). This is also in line with the apprehension principle of good graphical representation design, which states that graphics should be “accurately perceived to be accurately conceived” (Tversky, B., Morrison, J. B., & Betrancourt, 2002).

Clarity in graphical representation design leads to improved learning by increasing cognitive effectiveness of the diagrams, keeping the user’s motivation up and reducing fear of failure (Cennamo, 2016; Moreno, 2006). Research in Educational Technology R&D proposes to achieve clarity in graphical representation design by using simple and clear visual notation (Cennamo, 2016): the meaning of every object type, relational structure and visual variable used on the diagram should be salient and obvious. Therefore, we discuss a four practical guidelines to design transparent graphical representations.

Guideline 1: Choose semantically transparent and semiotically clear visual notation. First, the diagram needs to be able to syntactically represent all of the semantical constructs of the problem (Cox & Brna, 1995). Therefore, a diagram should be built with a visual notation which offers realistic depictions of its semantical constructs (i.e. a semantically transparent visual language) and which is also semiotically clear (i.e. assuring a 1:1 relationship between visual syntax and semantics) (Berenbach, 2004; Moody, 2009). Practically, the guideline of ‘first focus on content, then on appearance’ (Ambler, 2005) indicates that a diagram engineer should first decide on the different semantical constructs that are needed before picking a visual notation. This is supported by the congruence principle, which states that graphical formats should be chosen according to the content they need to present (Tversky, B., Morrison, J. B., & Betrancourt, 2002). The more semantically transparent the graphical symbols, the easier the graphical representation is to learn (Ormrod, 1999). For concrete information, a more visually expressive yet simple visual language like Oracle might work, while more advanced languages might only have a marginal effect. However, more abstract information might require more advanced visual notations like UML, which are more semantically expressive, but unfortunately less visually expressive (see Visualising - Principle of expressiveness below) and semantically transparent (Moody, 2009).

Guideline 2: Ensure clarity in structure. Second, the diagram should have clarity in structure. A clear structure (visual-spatial properties) is required to reflect the relationships
between diagram elements in a salient, semantically transparent way (Cox & Brna, 1995). Practically, the diagram should have: (1) an obvious and single ‘entry point’ (i.e. the point to start reading the diagram) (Berenbach, 2004), (2) verticality in hierarchal structures (Schuette & Rotthowe, 1998), (3) maximal edge and node orthogonality (Schuette & Rotthowe, 1998; Ware et al., 2002), (4) clear and continuous path flows (ref. principle of good continuity) and symmetry to make it easy to navigate through the diagram from start to finish (Ding & Mateti, 1990; Field et al., 1993; Colin Ware et al., 2002), (5) moderate length of paths to maintain understandability (Ware et al., 2002), (6) a minimal length of edges (lines) and arrows (Schuette & Rotthowe, 1998), (7) minimal bending of lines, as start- and ending point are less clear than those of straight lines (Ding & Mateti, 1990), and finally, (8) top-bottom ordering for high-to-lower levels (Ding & Mateti, 1990b)

Guideline 3: Be consistent in the use of visual syntax. Third, visual syntax should be consistently applied across the diagram to achieve syntactical clarity (Ambler, 2005). More in particular, literature in Information Systems & Computer Science prompts to be consistent in redundant coding like sizing, colouring, fonts and flow direction (see Ordering - Principle of direction of information Visualising - Principle of expressiveness) (Ambler, 2005; Ding & Mateti, 1990). Diagram engineers should also be consistent in the use of domain terminology and naming across the diagram (Purchase, 2002).

Guideline 4: Provide a diagram legend. Furthermore, not only is it important to be consistent in the use of visual syntax, it is also important to communicate the syntax properties of each semantical construct to the user, ideally by providing a legend with the diagram (Berenbach, 2004). Also, to increase usability and fasten lookup, choose a fixed place on the diagram (like a corner, or at the bottom) to locate the legend and place it there consistently across diagrams (Ambler, 2005).

Guidelines 5: Use concrete and explicit naming and labelling. Lastly, naming and labelling is an important part of a graphical representations (Irani & Ware, 2003) and can ease navigation between or within diagrams (Berenbach, 2004). When labelling, use concrete and explicit naming. Use verbs labels instead of nouns (Amar & Stasko, 2004; Ambler, 2005) and use lower-level aggregation naming for object features. Also, it is beneficial to clarify cause-effect relationships if there are any (Amar & Stasko, 2004).

Principle 10. Visualising - Principle of expressiveness

The principle of expressiveness states that diagram readability can be improved through the use of (planar and retinal) visual variables.
**Underlying theories.** While the number of elements on a diagram is limited by the working memory (Cowan, 2010; Miller, 1956), the sophistication, or visual expressiveness of the elements is not (Sweller et al., 1998). By using visual variables (e.g. colour, shape, texture), meaning can be added to a diagram without aggravating intrinsic cognitive load (i.e. cognitive processing), as visual variables are part of the extraneous load (perceptual processing phase). Also, the structural object perception theory states that in the first phase of human object recognition, graphical representations are scanned for fundamental visual variables to segment the image to facilitate information extraction (Irani & Ware, 2003). More attractive, i.e., visually expressive and aesthetic, diagrams are able to keep user motivation higher, leading to improve learning (Gaissmaier et al., 2012).

Bertin (1981) distinguished eight visual variables, which can be split into two groups: planar variables and “retinal variables” (Table 2). The eight variables can be applied on the level of an instance, line or area (Fout! Verwijzingsbron niet gevonden. in Appendix E) (Roberts, 2000).

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Power</th>
<th>Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planar variables</td>
<td>Horizontal position</td>
<td>Interval</td>
<td>10-15</td>
</tr>
<tr>
<td></td>
<td>Vertical position</td>
<td>Interval</td>
<td>10-15</td>
</tr>
<tr>
<td>Retinal variables</td>
<td>Size</td>
<td>Interval</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Brightness value</td>
<td>Ordinal</td>
<td>6-7</td>
</tr>
<tr>
<td></td>
<td>Colour</td>
<td>Nominal</td>
<td>7-10</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
<td>Nominal</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
<td>Nominal</td>
<td>30+</td>
</tr>
<tr>
<td></td>
<td>Orientation</td>
<td>Nominal</td>
<td>4</td>
</tr>
</tbody>
</table>

Various authors argue that engineers should use the full scope of visual variables to maximise visual diagram expressiveness (Ding & Mateti, 1990; Moody, 2009). The higher the level of visual expressiveness (i.e. lower degrees of visual freedom), the less room for interpretation or confusion from the side of the user (Westelinck et al., 2005). The latter is exactly the strength of the graphical representations over textual formats: improving the quality of communication by improving understandability of information through visual expressiveness (Goldman, 2003). Yet, many SE notations still only make use of one or two retinal variables to encode information (“primary notation”) and leave a broad range of visual variables ‘open’ to be chosen by the designer (“secondary notation”). For example, ER diagrams and DFDs, two leading SE notations, only make use of shape as a retinal variable. This means they only have a visual expressiveness score of one out of a scale of eight, thus
leaving a lot of room to improve expressiveness. Suboptimal visual expressiveness and often inconsistent visual encoding across diagrams results in erroneous interpretation, ultimately nullifying the advantage of graphical representations over textual formats (Moody, 2009; Westelinck et al., 2005).

Following the congruence principle (i.e. “form follows content”), the semantics should determine the choices in secondary visual notation (i.e. the visual variables not specified by the chosen visual language) made by the designer (Ambler, 2005; Ding & Mateti, 1990; Tversky, Morrison, & Betrancourt, 2002). These ‘informal’ syntax choices will largely determine the readability of the diagram and should be the result of proper design rationale. For example, by drawing arrows, the engineer will determine reading direction, through the indication of start and end points (Ding & Mateti, 1990). Also, by opting for a tree diagram, the designer will signal the hierarchal nature of the information, while a list of elements will not do so (Ding & Mateti, 1990). In what follows, we discuss the use of Bertin’s visual variables from both a theoretical and practical standpoint to optimise diagram expressiveness.

Guideline 1: Make use of sizing effects. Diagram elements should be sized in correspondence to the total diagram size, the number and types of symbols, sizes of other symbols, symmetry and regularity requirements (Ding & Mateti, 1990). Too large or too small elements make the diagram difficult to read. Furthermore, sizing should be consistent in order to maintain readability and avoid unwanted visual noise effects (users extracting meaning from unpurposely differing sizes) (Ding & Mateti, 1990).

Guideline 2: Use differences in brightness. The brightness value reflects the colour intensity or ‘value’ of a diagram element (Bertin, 1981), which allows for revealing order in information (Roth & Mattis, 1990). In contrast to the retinal variable colour, an item’s brightness value will remain visible when a graphical diagram is viewed in black-and-white (no loss of information), and can thus be used to encode non-redundant information.

While brightness is especially recommended to express ordinal positioning of within diagram elements, it will have weak results applied on nominal object types (Pilgrim, 2003). In Earth Science, brightness value is used to indicate specific type of clouds to ease satellite image interpretation (Ogao & Kraak, 2002).

Guideline 3: Make use of colour to differentiate between element types. Learning and Instruction literature promotes the use of colour to differ between different exemplar types (Ormrod, 1999), while Data Science frequently uses colour to mark patterns, indicate icon
attributes, and increase diagram expressiveness (Koshman, 2010). Computer Science guidelines prompt to use colour to improve readability of diagrams (Ambler, 2005). In a Java case study, Coad, Lefebvre and DeLuca (1999) describe the use of colour on UML to improve understandability of diagrams and support software and system development. In the latter, colour is used to indicate development priority, preferred implementation languages for certain classes, or to describe the target platform for a specific module (Ambler, 2005).

Guideline 4: Make use of texture to differentiate between nominally measured element types. Texture can be used to eliminate the need for labelling and therefore decrease the cognitive load of a graphical representation (Irani & Ware, 2003). While for interval and ordinal scope of measurement, texture has found to be only marginally effective, it has found to be a good and effective retinal variable for objects measured on a nominal scale (Pilgrim, 2003).

Guideline 5: Consistently use of shape to difference between element types. As different shapes have different meanings for users from different fields and with diverse prior knowledge, the use of more ‘standard shapes’ like circles, rectangles, triangles, ellipses and diamonds (squares) are recommended unless specific (domain-related) shapes are required (Ding & Mateti, 1990). However, it is important to note that research in Cognitive Psychology has found that curved shapes like circles and ellipses are more effective than rectangular forms and are preferred by users (Bar & Neta, 2006; Irani & Ware, 2003; Marr & Nishihara, 1978). The use of shapes should be consistent within and across diagrams (Ding & Mateti, 1990). In Information Systems & Computer Science, rectangles are mostly used for tasks, diamonds for decision points, and circles to represent nodes on different levels or phases. The variable shape is one of the most frequently used, and often the sole retinal variable in many (SE) notations. However, it is found to be the least expressive one of all visual variables because it can only represent information on a nominal scale and it is cognitively the least effective (Lohse, Min, & Olson, 1995; Moody, 2009). Therefore, it is strongly recommended to make use of other visual variables as well to increase the expressive capacity of the graphical representation.

Guideline 6: Be consistent in element orientation. In line with the principle of direction of information, experiments in the field of Vision Research have shown that inconsistent element orientation relative to the general path orientation (i.e. the general reading direction of the diagram) has a negative impact on user performance (Field et al., 1993). In other words, it is beneficial to be consistent with orientation within and across diagrams to increase readability. One remark to make here is to keep the target audience or diagram task in mind: some users of software applications might expect a particular orientation of
elements (e.g. some users are used to vertically oriented one-dimensional arrays, while others to horizontal arrays) (Ding & Mateti, 1990). Additionally, orientation can be used to segment diagram elements into groups (Field et al., 1993): following Gestalt Theory, elements with the same orientation seem to exert some kind of ‘common fate’ and are therefore perceived as a group of elements (Koffka, 2013).

**Guideline 7: Horizontal and vertical position.** Horizontal and vertical positioning of elements can be used to build hierarchical structures and reflect inter-element relationships. The positioning of elements also determines the symmetry of the diagram, which can improve readability (Schuette & Rotthowe, 1998; Colin Ware et al., 2002).

**Guideline 8: Use intensity of visual variables to signal criticality of diagram elements.** From research in *Cognitive Psychology*, it seems that all the information in the working memory can potentially be made available to an individual’s consciousness, but what a user is actually conscious of at a certain moment, is determined by attention (Rowlatt, 2008). Focussing attention has positive effects on information transfer, knowledge compilation and generating relational statements (Mautone & Mayer, 2007; McAndrew, 1983; Paas et al., 2004; Van Merriënboer & Kester, 2005). Visual variables can be used to signal the importance of diagram elements and reduce search (Van Merriënboer & Kester, 2005). *Learning & Instruction* literature argues that intensity of visual variables can be used to guide attention of the user (Ormrod, 1999). The intensity of a diagram element is determined by its size and colour. The larger the graphical element, the higher the perceived importance by the user, and the more attention it gets (Ormrod, 1999). Colour can make any element or group on the diagram pop-out, drawing attention to it and therefore enhancing its processing (Ormrod, 1999). For example, the colour red is often used to indidate critical diagram parts.

**Guideline 9: Use novelty and unexpectedness to increase expressiveness.** Novelty and unexpectedness can create attention effects as well, and therefore increase diagram expressiveness. When an element, image or icon is initially unexpected within the context, it creates a surprise effect, increasing attention and retention of that element (Ormrod, 1999). For example, formatting information in unexpected, hard-to-read fonts can enhance attention and thereby improve recall (Diemand-Yauman, Oppenheimer & Vaughan, 2011).

**Guideline 10: Use the principle of expressiveness with moderation.** Although there are many visual variables and techniques that have found to improve diagram expressiveness (Armbruster, 1984; Lorch, Lorch & Inman, 1993; Reynolds & Shirey, 1988), the right balance has to be found. Too much focussing will lead to *tunnel vision* effects, yet too little focus will lead to *splatter vision* (i.e. peripheral vision without any focus) (Fiol & Huff, 1992). While
multiple authors recommend using the full range of visual variables, *Learning & Instruction* guidelines warn for overloading the diagram design with retinal variables, as this can have negative effects as well. Research in the field found that novices get easily distracted by perceptually prominent instances and consequently miss the fundamental semantics behind the syntax (Lowe, 2003). Also, embellishments extraneous to perceptual or cognitive processing increase cognitive load and may decrease diagram readability (Schnotz, 2013). Some research even argues that instead of trying to maximise visual expressiveness, graphical representation design should rather aim to minimise visual expressiveness by maximising construct specificity and decreasing the need for retinal variables (Reimann, 2003). We recommend that graphical representation designers systematically make the trade-off between visual expressiveness and diagram readability depending on the particular diagramming goal.

### 3.2.3. Integrating Principles

So far, we have presented fundamental and practical graphical representation design principles to maximise diagram readability. In what follows, two *integrating principles* are presented to resolve conflicts between principles and evaluate design choices for different end-tasks and end-audiences.

**Principle 11. Principle of Differentiation**

| The principle of differentiation states that diagram readability can be improved by differentiating diagram design for user characteristics, task characteristics and semantical requirements of the graphical representation. |

#### Underlying theory

Different visual-spatial characteristics are appropriate in different contexts (Cox, 1999). The three context factors that influence graphical representation appropriateness (and thus effectiveness) are *user characteristics* (a combination of prior knowledge, cognitive abilities, cognitive style), *task characteristics* (fit for purpose) and *semantical requirements* (the ability of the visual syntax to convey the semantics) (Ainsworth, 2006; Cox, 1999). When these factors are known, graphical representation design can be optimised correspondingly, prioritising some graphical representation principles and strategies over others (Ainsworth, 2006). In what follows, we discuss considerations a diagram engineer should make when differentiating for target audience characteristics, tasks characteristics and semantical requirements.

**Guideline 1: Use graphical representations for a novice target-audiences.** Experiments in *Computer Science* found that novices are better off with graphical representations than
with text, while for experts no difference in performance was found (ChanLin, 2001; Reimann, 2003).

**Guideline 2: Use gradual content build-up for novice target-audiences.** Second, the learning of complex information can be an intensive and demanding process, accommodating the graphical representation for the learners’ prior knowledge level is important to keep motivation high and fear of failure low (Cennamo, 2016; Van Merriënboer & Sweller, 2005). *Learning & Instruction* researchers argue that knowing user prior knowledge levels is crucial to determine the right level of element interactivity and refinement to communicate information (Sweller et al., 1998; Van Merrienboer, 1997). For example, novices benefit from using fading techniques to gradually build up information provision. Practically, for novices, it might be a good idea to start with concrete concepts and examples before moving towards a more general, abstract level (i.e. an *inductive-expository strategy*) (Merriënboer, Clark, & Croock, 2002). For experts however, information can be represented at a high level of abstraction and move from there towards the learning tasks (i.e., a *deductive strategy*) (Merriënboer et al., 2002). While a deductive strategy is effective for experts, novices might experience learning difficulties with this representational strategy (Merriënboer et al., 2002).

**Guideline 2: Do not represent too much details for expert target-audiences.** The *expertise reversal effect* describes how redundant information on representations (which might be needed for novices) increases cognitive load for experts and decreases performance (Yeung et al., 1999). Therefore, when the target-audience is one of experts, graphical representations should not be overly detailed (McNamara, Kintsch, Songer, & Kintsch, 1996; Sweller et al., 1998). Additionally, research found that reduced information provision prompts the advanced learner to process more deeply and actively, resulting in positive learning effects (McNamara et al., 1996; Yeung et al., 1999).

**Guideline 3: Make use of audience familiarity with certain visual syntax and visual languages.** Users from different fields have different expectations and conventions regarding lay-out, notation, naming, meta-models etc. (Becker et al., 2000). An illustration of this is that programmers generally like to work with boxes or rectangular shapes, researchers in the field of *Graphics* prefer to work with circular shapes (Ding & Mateti, 1990). However, the most conventional or familiar representation mode is not always the best choice. The engineer should make a trade-off between the cost of less familiarity with the
benefit of higher computational offloading effects of a more unknown visualisation technique (Peebles & Cheng, 2003).

**Guideline 4: Prefer graphical representations for high performing students and textual formats for lower performing students.** Differences in learning and cognitive abilities of the user should be taken into consideration as well (Cox, Stenning, & Oberlander, 1994). Research in *Cognitive Science* has found that high performing students profit from graphically formatted instruction, while lower performing students profit from textual instruction (Cox et al., 1994).

**Guideline 5: Differentiate for cognitive styles of users.** The effectiveness of reasoning with diagrams is affected by the instructional preferences (i.e. cognitive style) of the user (Cox, 1996). Depending on where individuals are situated along the visualiser-verbaliser (VV) dimension of cognitive style, they differ in the way they reason with diagrams and are able to distract information from different formats (Cox & Brna, 1995). A graphical representation designer should differentiate his or her design for the cognitive style of the target audience.

**Guideline 6: Differentiate for task type.** Learner performance is improved when the structure and representation format of the learning material is in correspondence with the task requirements (ref. cognitive fit theory) (Ainsworth, 2006; Claes, Gailly, & Poels, 2013; Gilmore & Green, 1984). Depending on the diagramming purpose, different graphical design principles will be prioritised. For example, where diagrams built for documentation or specification purposes should aim for exhaustive and precise visualisation modes, diagrams built for communication purposes should rather focus on visualisations which are clear and highly understandable.

**Guideline 7: Differentiate for semantical requirements.** Diagram engineers should evaluate visualisation techniques for their ability to represent the semantics of the domain and to facilitate understanding of the domain (Gemino & Wand, 2003). Depending on the specific semantical requirements of the information to be presented, the designer should differentiate his or her visualisation strategy. Studies in the field of *Learning & Instruction* find that depictive formats (e.g. flowcharts) are better suited to represent abstract and complex information, while descriptive formats (like verbal or textual format) are better for more concrete, lower-complexity information (Cox, 1999; Schnotz, 2013).

**Illustration of differentiation techniques in practice.** A new emerging field that is particularly strong in general format differentiation is *Adaptive Learning & Instruction*. 
Adaptive learning software takes learner’s ongoing level of expertise and prior knowledge into account and dynamically adjusts its content provision and its content provision format accordingly (Van Merriënboer & Sweller, 2005). This way, graphical representations are optimally differentiated for the user and attain higher levels of effectiveness.

To test graphical representations for their level of adaptation and differentiation in a specific case, research in Learning & Instruction by Gemino & Wand (2003) argues to combine both a theoretical approach (for testing semantical representation power) and empirical approach (user-based approach to test user characteristics). This way, fit-for-purpose of the graphical representation design is assessed.

**Principle 12. Principle of Balancing**

| The principle of balancing argues that when a graphical representation must accommodate multiple end-users and end-tasks, graphical representation design should be balanced by (1) using multi-media techniques and by (2) assigning weight of relative importance to the different fundamental and practical principles of graphical representation design. |

**Underlying theory.** In some cases, graphical representations cannot be customised for target audience or tasks (Principle of Differentiation) because they are used by a diverse range of end-users and for various end-tasks. Graphical representation design then needs formal techniques to balance its design to optimally accommodate for these different purposes.

**Guideline 1: Using multimedia to accommodate multiple end-users and end-tasks.** The first way to balance the graphical representation design for diverse end-users or multiple purposes, is to make use of the multimedia effect (Moreno & Mayer, 1999). Offering multiple representational formats allows the user to choose the best one for the task at hand and adapt strategy (Ainsworth, 2006; Dunn & Dunn, 1993). Additionally, a single representation can be ‘multimedia’ through combining depictive formatting with text. Mayer’s multimedia effect states that learners learn more from the combination of words and graphics together than from words alone (Moreno & Mayer, 1999; Van Merriënboer & Kester, 2005). However, research on the effectiveness of the multimedia effect has resulted in contradicting results. Various studies found positive results (Cox & Brna, 1995; R. E. Mayer & Sims, 1994), but just as many did not find those benefits (Ainsworth, 2006; Chandler & Sweller, 1992). For example, studies on analogical reasoning show that learners’ understanding improves from comparing different information sources (Gentner & Markman, 1997). However, various
studies have found that users have difficulties with integrating information from multiple sources and tend to treat the multiple representations in isolation (Ainsworth, 2006).

In his *integrated model of text and image comprehension*, Schnotz (2013) argues that there are conditional principles for the use of multimedia. He argues that (1) combining diagrams and text should only be done when learners have low prior knowledge but enough cognitive abilities to process both the depictive and descriptive information, (2) depictive and descriptive formats should not be combined if one or the other is redundant, (3) pictures should only be combined with text when their semantical relation is clear, (4) text and graphical representations should be presented in close spatial proximity. Multimedia can thus be used to improve learning from graphical representations, but only under the stated conditions.

**Guideline 2: Balance the fundamental and practical graphical representation design principles to accommodate multiple end-users and end-tasks.** It is clear that when designing graphical representations according to the design principles, the engineer might encounter conflicts between principles (e.g., maximising diagram minimality versus maximising diagram expressiveness). Also on a lower level, conflicts between *design rules* (i.e., guidelines within a principle) might have to be resolved (Figure 18).

![Figure 18: Illustration of conflicting rules. (a) satisfies the rule of 'minimise number of crossings', while (b) satisfies the rule of 'minimise total path length'. Source: Ding & Mateti (1990)](image)

Depending on the context, not all principles or rules are equally important (Ding & Mateti, 1990). Graphical representation design is not about insisting on extremes (Purchase, 2002), but rather about making purposeful trade-offs between various low-level design objectives (e.g., clarity, visual expressiveness, completeness, level of refinement etc.) in order to achieve the high-level diagram objective (e.g., effective information transfer to learners with low prior knowledge). For example, Long (2002) discusses an often made trade-off in the field of SE, where the benefits of enhanced control (e.g., using more control constructs on a diagram) are balanced against the benefits of increased data-triggered execution (e.g., by using less control constructs on a diagram). Giving higher priority to one or the other will result in different graphical design strategies (Appendix C).
Because of this frequent multiplicity in end-users and end-tasks of graphical representations, it is increasingly relevant and important to evaluate the quality of graphical representations from different viewpoints (Becker et al., 2000). These different viewpoints are embodied within the graphical representation design principles that we have discussed above. Each principle represents a different component and perspective on diagram effectiveness and should be given consideration. While it is never beneficial to completely neglect a graphical design principle, appropriate design rationale can allocate higher priority to certain principles without completely disregarding the others (Ding & Mateti, 1990).

This challenging balancing exercise has a high impact on the quality of the graphical representation, yet little literature exists on this topic. The lack of clear balancing conventions and uniformity in design processes leads engineers to build and balance graphical representations according to their preferences or habit, leading to often suboptimal and ineffective graphical representations (Berenbach, 2004). Therefore, we propose a formal solution for balancing the design principles and rules. We propose to measure the effectiveness of graphical representations by continuous aesthetic metrics. Evaluating graphical designs on a continuous spectrum allows to assess the extent to which they adhere to certain graphical design principles, rather than assessing binary (yes/no) conformance (Purchase, 2002).

A paper by Purchase (2002) proposes seven objectively defined metrics (Table 3) to quantify the extent of aesthetic conformance of a graphical representation.

<table>
<thead>
<tr>
<th>Nr</th>
<th>Rule</th>
<th>Graphical Design Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minimise number of crossings</td>
<td>Principle of Distinction</td>
</tr>
<tr>
<td>2</td>
<td>Maximise edge orthogonality</td>
<td>Principle of Distinction</td>
</tr>
<tr>
<td>3</td>
<td>Maximise node orthogonality</td>
<td>Principle of Distinction</td>
</tr>
<tr>
<td>4</td>
<td>Minimise edge bends</td>
<td>Principle of Distinction</td>
</tr>
<tr>
<td>5</td>
<td>Maximising minimum angle</td>
<td>Principle of Distinction</td>
</tr>
<tr>
<td>6</td>
<td>Maximise symmetry</td>
<td>Principle of Expressiveness</td>
</tr>
<tr>
<td>7</td>
<td>Rule of consistent flow direction</td>
<td>Principle of Clarity</td>
</tr>
</tbody>
</table>

The metrics are formally quantified through a mathematical algorithm which ultimately outputs scores between 0 (zero conformance) and 1 (full conformance) for each rule. This
way, Purchase measures the aesthetic conformance of graphical representations from different perspectives. An illustration is given in Appendix F.

We will use this evaluation technique and algorithm developed by Purchase as a basis for formally balancing graphical representation design. We propose to take a similar approach, but rather one that takes all of the Fundamental and Practical Graphical Representation Design Principles into account in an attempt to be more exhaustive. Through assessing the underlying lower-level rules and guidelines of each principle, quality in graphical representation design can be measured. This evaluation technique can also be used to compare different graphical representations for the same information for certain purposes. Depending on the target audience and intended end-tasks, weights can be assigned to the different rules to indicate and formalise relative importance (Ding & Mateti, 1990). For example, on Figure 18, the rule of ‘minimise crossings’ can be given a higher impact than the rule of ‘minimising total path length’, so that representation mode (a) will be preferred. It is important to remark here that the weighting of rules will then add a subjective notion to this so far formal and objective evaluation technique.
4. Conclusion and discussion

4.1. The physics of graphical representations: a theory for graphical representation design

In an attempt to create a scientific basis for graphical representation design, twelve principles for qualitative graphical representation design have been defined. The principles are divided into three groups: the Fundamental Principles, the Practical Principles and the Integrating principles.

With these twelve principles, we aim to provide an initial scientific basis for graphical representation design.

4.1.1. The Fundamental Principles

The three Fundamental Principles provide the foundation for and reasoning behind the seven Practical Principles. By using theories from Cognitive Science, the Principle of Diagram Acquisition emphasises the need for diagram readability, which leads to the need for minimality (Principle of Minimality) and the need for complexity management (Principle of Complexity Management).

4.1.2. The Practical Principles

The Practical Principles indicate how graphical representation design can achieve conformance to the Fundamental Principles and thus achieve diagram effectiveness. The Practical Principles are divided into three categories: the Grouping Principles (Chunking and Modularisation), the Ordering Principles (Direction of Information, Internal and External Linkage) and the Visualising Principles (Distinction, Clarity and Expressiveness). More concrete and practical (but yet research-based) rules and guidelines are presented for each Practical Principle to guide the engineer in the design process.

4.1.3. The Integrating Principles

Finally, the Integrating Principles indicate how to balance (Principle of Balancing) and resolve principle conflicts depending on the target audience and intended end-tasks (Principle of Differentiation).

4.2. Practical evaluation and significance

The principles of this paper can be used to:
- Design graphical representations: Having explicit design principles helps to create uniformity in the process of graphical representation design and to formalise and guide design decisions.

- Compare graphical representation designs: Different diagram designs can be compared by analysing how well each of them performs against certain design principles.

- Evaluate and improve graphical representation designs: The principles can be used for formalising quality assurance of graphical representation design. Through the use of the quality evaluation technique discussed under the Principle 12, the impact of design decisions can be ‘quantified’. The metric score system allows a diagram engineer to see what the relative impact is of particular design decisions on the conformance to the different diagram design principles. This way, basic errors against the principles of graphical representation design can be avoided (Berenbach, 2004).

Various stakeholders benefit from the application of the principles. Graphical representations are used to communicate to both novice and expert users, and between them. Novices (e.g., customers, top-level management) will be able to better understand the represented information, leading to higher efficiency and less errors. Experts will be able to better collaborate through standardised design processes and formats (Scaife & Rogers, 1996), and also enhance maintainability5 of diagrams (Nelson et al., 2012; Zugal, 2013). Finally, the consequences of novice-expert differences (e.g., communication problems, expertise reversal effects, reading difficulties for novices) are minimised.

4.3. Theoretical evaluation and significance

4.3.1. Prior research and current problem situation

Existing research in the field of Information Systems & Computer Science on the quality in graphical representations has been mostly centred around the semantical and ontological analysis of graphical representations (the right-hand side of Figure 1). Illustrations of this are respectively the LSS framework of Lindland, Sindre and Sølvberg (1994) and the Bunge–Wand–Weber (BWW) ontology (1990). The role of visual syntax has been systematically undervalued and efforts to provide graphical representations with syntactical design rationale have been largely absent (Chen, 2005; Moody, 2009; Rogers & Scaife, 1998). As a result, many engineers are ineffectively designing diagrams according to personal preferences

---

5 The maintainability of a diagram refers to the ease with which a diagram can be improved, extended or used later on by other experts.
(Rogers & Scaife, 1998). This lack of directedness prevents the field of Graphical Representation Design to grow and make scientific progress.

4.3.2. Scoping this paper on the graphical representation anatomy

The void of research about the syntax of graphical representations has been recently partly addressed by Moody with his Physics of Notations (Moody, 2009). Moody’s work focusses exclusively on quality in visual notations (top-left quadrant in Figure 1), yet does not say anything about how to use these visual notations in an effective manner. This paper complements Moody’s syntactical analysis on language level with principles for syntactical analysis on a sentence (diagram) level (bottom-left quadrant on Figure 1).

4.4. Limitations and further research

As mentioned, this work does not go into how to define semantical constructs for diagrams (top-left quadrant on Figure 1) or how to use them on a sentence level (bottom-left quadrant on Figure 1). Additionally, this paper also does not dictate what type of graphical representation (graphs, tables, Venn diagrams, decision tree, flow chart, etc.) one should use for defined types of scenarios. It only hints towards what type of information matches best with what type of visualisation mode, or what type of visualisation strategy matches best with what type of target audience. No exhaustive overview or evaluation of the existing graphical representation techniques and strategies is provided.

This paper takes a user-perspective on graphical representation design and focusses on the decoding side, yet, the encoding side, i.e. the engineer-perspective, is also important to consider. Ease of the designing and maintainability of the representations are also measures for quality in graphical representation design and are valuable to consider (Evitts, 2000; Viyović, Maksimović, & Perišić, 2014).

Last, while the literature study is assumed to exhaustively cover the problem space of the use of visual syntax on a sentence level, it is likely that not all existing descriptive theories and prescriptive guidelines are covered in this paper.

We encourage future research to build further on a scientific basis for graphical representation design and take over where this paper left off.

4.5. Wider significance

While we have often referred to the field of Information Systems & Computer Science for both problem identification (determining weak spots) and solution identification (taking over research-based best practices and conventions), the graphical representation design principles defined in this paper also apply wide outside this field. Examples throughout the
References


Dunn, R. S., & Dunn, K. J. (1993). *Teaching secondary students through their individual learning styles: Practical approaches for grades 7-12*. Prentice Hall.


Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and
temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001


https://doi.org/10.1109/INES.2014.6909375

https://doi.org/10.1057/palgrave.ejis.3000686

https://doi.org/10.1109/32.360316


https://doi.org/10.1057/palgrave.ivs.9500013

https://doi.org/10.5381/jot.2003.2.6.r1


Wheildon, C., & Heard, G. (2005). *Type & layout: Are you communicating or just making pretty shapes? Are you communicating or just making pretty shapes? How typography and design can get you message across—or get in the way.*


https://doi.org/10.1080/00220979909598353


## Appendix A

<table>
<thead>
<tr>
<th>Prescriptive Guidelines &amp; Descriptive Theories</th>
<th>C</th>
<th>R</th>
<th>V</th>
<th>T</th>
<th>Principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Load Theory</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Focus on diagram acquisition</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>The magical number 7</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Cowan's model of attention and memory</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Split attention effect</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Communication is the purpose of graphical representations</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Computational offloading</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Visual fitting</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Redundancy avoidance</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Coherence principle</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Preference for reduced and simplified information</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Determinism guideline</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>The magical number 7</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Reduce semantic complexity</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Introduce symbol deficit</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>&quot;Only show what you have to show&quot;</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Prevent overlap</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Apprehension principle</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Cognitive load theory</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Higher levels of motivation lead to improved learning</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Preference for single-page diagrams</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Preference for simplified information delivery</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Principle of focus on diagram acquisition</td>
</tr>
<tr>
<td>Neighbourhoods guideline (spatial proximity)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Chunking - Principle of grouping</td>
</tr>
<tr>
<td>Modularization (categories)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Chunking - Principle of modularisation</td>
</tr>
<tr>
<td>Gestalt theory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chunking - Principle of modularisation</td>
</tr>
<tr>
<td>Schema theory</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Nesting</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Fading principle</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Maximizing consistent flow direction</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Have verticality of hierarchical structures</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Maximise orthogonality of diagram</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Ordering - Principle of direction of Information</td>
</tr>
<tr>
<td>Prior knowledge activation</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Repleteness principle</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Strengthening</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Semantic transparency</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Reusable patterns: strengthening</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Make use of existing knowledge</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Dimensional sequences as determiners of cognitive organization</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Weak coupling</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Connect and integrate abstract and concrete representations of concepts</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Integration</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Split attention principle</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Stretch text provision</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Ordering - Principle of internal and external linkage</td>
</tr>
<tr>
<td>Core versus detail</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Spatial separation: avoid close lines</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Uniform density of nodes</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Maximise edge and node orthogonality</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Minimise number of crossings</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Distinctive shapes</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Make use of redundant coding</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Minimum angle</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
<tr>
<td>Avoid diagonal or bending lines</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>Focusing - Principle of distinction</td>
</tr>
</tbody>
</table>

### Legend OR Dimension
- C: Providing content
- R: Showing relationships
- V: Visualising content & relationships
- T: Timely presenting content & relationships
<table>
<thead>
<tr>
<th>Principle of semiotic clarity</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pragnanz principle</td>
<td>x</td>
</tr>
<tr>
<td>Transparency in use of symbols</td>
<td>x</td>
</tr>
<tr>
<td>Congruence principle: appearance follows content</td>
<td>x</td>
</tr>
<tr>
<td>Consistency guidelines (font, sizing, colour, naming)</td>
<td>x</td>
</tr>
<tr>
<td>Optimise length of paths</td>
<td>x</td>
</tr>
<tr>
<td>Structure vertically and maximise orthogonality</td>
<td>x x</td>
</tr>
<tr>
<td>Explicit labeling and naming</td>
<td>x</td>
</tr>
<tr>
<td>Principle of Good Continuity</td>
<td>x</td>
</tr>
<tr>
<td>Consistency in diagram legend placement</td>
<td>x x</td>
</tr>
<tr>
<td>Single and clear entry point</td>
<td>x x</td>
</tr>
<tr>
<td>Length of path and edges</td>
<td>x</td>
</tr>
<tr>
<td>Sophistication of elements</td>
<td>x</td>
</tr>
<tr>
<td>Visual variables of Berlin</td>
<td>x</td>
</tr>
<tr>
<td>Guideline of distinctive appearance</td>
<td>x</td>
</tr>
<tr>
<td>Congruence principle: appearance follows content</td>
<td>x x</td>
</tr>
<tr>
<td>Structural object perception theory</td>
<td>x</td>
</tr>
<tr>
<td>Gestalt Theory Laws</td>
<td>x</td>
</tr>
<tr>
<td>Disfluency effect: ugly fonts</td>
<td>x</td>
</tr>
<tr>
<td>Provide a notation legend</td>
<td>x</td>
</tr>
<tr>
<td>Maximise diagram symmetry</td>
<td>x</td>
</tr>
<tr>
<td>Multimedia theory</td>
<td>x</td>
</tr>
<tr>
<td>Readability</td>
<td>x</td>
</tr>
<tr>
<td>Use what is unexpected to get attention</td>
<td>x x</td>
</tr>
<tr>
<td>Visual emphasis (signalling principle)</td>
<td>x x</td>
</tr>
<tr>
<td>Guideline of maximal specificity</td>
<td>x</td>
</tr>
<tr>
<td>Attention</td>
<td>x x</td>
</tr>
<tr>
<td>Cognitive flexibility theory</td>
<td>x</td>
</tr>
<tr>
<td>Splitter vision versus tunnel vision effect</td>
<td>x</td>
</tr>
<tr>
<td>Split attention principle</td>
<td>x</td>
</tr>
<tr>
<td>Target group analysis</td>
<td>x</td>
</tr>
<tr>
<td>Adaptive learning</td>
<td>x x</td>
</tr>
<tr>
<td>Cognitive fit theory</td>
<td>x x</td>
</tr>
<tr>
<td>Visualiser-verbalsiser (Vv) dimension of cognitive style</td>
<td>x</td>
</tr>
<tr>
<td>Training wheels principle</td>
<td>x</td>
</tr>
<tr>
<td>Inductive-expository strategy</td>
<td>x</td>
</tr>
<tr>
<td>Deductive strategy</td>
<td>x</td>
</tr>
<tr>
<td>Fading principle</td>
<td>x x</td>
</tr>
<tr>
<td>Expertise reversal effects</td>
<td>x x</td>
</tr>
</tbody>
</table>

### Visualising - Principle of clarity

<table>
<thead>
<tr>
<th>Principle of differentation</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principle of balancing</td>
<td>x</td>
</tr>
</tbody>
</table>

### Integrating principles

<table>
<thead>
<tr>
<th>Principle of differentation</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principle of balancing</td>
<td>x</td>
</tr>
</tbody>
</table>

| Multimedia effects           | x x |
| Integrated model of text and image comprehension | x x |
| Conflicting rules            | x |
| Evaluate GR quality from different viewpoints | x |
| Congruence principle: appearance follows content | x x |
| The control enablement and data triggering spectrum | x |
| Aesthetic metrics for evaluating GR | x |
Appendix B

Figure 19: Illustration of the Gestalt law of good continuation. The continuity of the branches is easily identified although the grid disrupts the image. Source: Field et al. (1993)
Figure 20: Comparing graphical representations on the Behavior Characteristics Spectrum: from FFBDs (Function Flow Block Diagrams) to DFDs (Data Flow Diagrams). Source: Long (2002)
Figure 21: Illustration of spatially integrated text and picture. Spatial proximity did not improve learning any further when text was segmented and pictures were labelled. Source: Florax & Ploetzner (2010)

Figure 22: Illustration of learning material with segmented text and labelled picture. Source: Florax & Ploetzner (2010)
Appendix E

Table 4: The retinal variables apply for points, lines and areas. Source: Roberts (2000)

<table>
<thead>
<tr>
<th></th>
<th>Point</th>
<th>Line</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td><img src="image1" alt="Size Point" /></td>
<td><img src="image2" alt="Size Line" /></td>
<td><img src="image3" alt="Size Area" /></td>
</tr>
<tr>
<td>Value</td>
<td><img src="image4" alt="Value Point" /></td>
<td><img src="image5" alt="Value Line" /></td>
<td><img src="image6" alt="Value Area" /></td>
</tr>
<tr>
<td>Colour</td>
<td><img src="image7" alt="Colour Point" /></td>
<td><img src="image8" alt="Colour Line" /></td>
<td><img src="image9" alt="Colour Area" /></td>
</tr>
<tr>
<td>Texture</td>
<td><img src="image10" alt="Texture Point" /></td>
<td><img src="image11" alt="Texture Line" /></td>
<td><img src="image12" alt="Texture Area" /></td>
</tr>
<tr>
<td>Shape</td>
<td><img src="image13" alt="Shape Point" /></td>
<td><img src="image14" alt="Shape Line" /></td>
<td><img src="image15" alt="Shape Area" /></td>
</tr>
<tr>
<td>Orientation</td>
<td><img src="image16" alt="Orientation Point" /></td>
<td><img src="image17" alt="Orientation Line" /></td>
<td><img src="image18" alt="Orientation Area" /></td>
</tr>
</tbody>
</table>
Table 5: Source: Illustration of the application of the seven aesthetic metrics of Purchase. Purchase (2002)

<table>
<thead>
<tr>
<th></th>
<th>$N_c$</th>
<th>$N_b$</th>
<th>$N_s$</th>
<th>$N_m$</th>
<th>$N_{eo}$</th>
<th>$N_{no}$</th>
<th>$R_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRCLE-4-ITS</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>COMPLETE-4</td>
<td>0.67</td>
<td>1</td>
<td>1</td>
<td>0.37</td>
<td>0.33</td>
<td>0.57</td>
<td>0.67</td>
</tr>
<tr>
<td>CIRCLE-8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.75</td>
<td>0.5</td>
<td>0.04</td>
<td>0.5</td>
</tr>
<tr>
<td>FIBONACCI-4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.78</td>
<td>0.57</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>BINARY-4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.77</td>
<td>0.34</td>
<td>0.01</td>
<td>0</td>
</tr>
</tbody>
</table>

Legend:

- $N_c$: Metric for number of crossings
- $N_b$: Metric for number of bends
- $N_s$: Metric for symmetry level
- $N_m$: Metric for minimum angle between edges leaving a node
- $N_{eo}$: Metric for edge orthogonality
- $N_{no}$: Metric for node orthogonality
- $N_r$: Metric for consistent flow direction