1 2 9 0

UNIVERSIDADE D
COIMBRA

João Miguel Andrade Proença da Cunha

Generation of Concept Representative Symbols: Towards Visual Conceptual Blending

Doctoral thesis submitted in partial fulfilment of the Doctoral Program in Information Science and Technology supervised by Professor Fernando Jorge Penousal Martins Machado and Professor Pedro José Mendes Martins, and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University

May 2022

1 2 9 0

UNIVERSIDADE D
COIMBRA

João Miguel Andrade Proença da Cunha

Generation of Concept Representative Symbols: Towards Visual Conceptual Blending

Tese no âmbito do Programa de Doutoramento em Ciências e Tecnologias da Informação orientada pelo Professor Doutor Fernando Jorge Penousal Martins Machado e pelo Professor Doutor Pedro José Mendes Martins, e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia

Maio de 2022

Generation of Concept-Representative Symbols:
Towards Visual Conceptual Blending

Supervised by Associate Professor Fernando Jorge Penousal Martins Machado and Assistant Professor Pedro José Mendes Martins

Examiners: Prof. Lydia B. Chilton, Prof. Pablo Gervás, Prof. Geraint A. Wiggins and Prof. Bernardete Ribeiro. Presided by Prof. Ernesto Costa

This work was supported by Fundação para a Ciência e a Tecnologia, under the grant SFRH/BD/120905/2016

FCT Fundação para a Ciência e a Tecnologia

Co-financed by:

POCH PROGRAMA OPERACIONAL CAPITAL HUMANO
CENTRO 2020
PORTUGAL 2020
UNIÃO EUROPEIA Fundo Social Europeu

To my avó Lindita and avô Mário,
who gave me strength to make this work come to fruition.

He constantly asking how he could be of help.

She constantly asking when my examination would be,
worried that she would not be present.

It is my belief that, somehow,
they are watching from the front row.

ABSTRACT

The visual representation of concepts has been the focus of multiple studies throughout history and is considered to be behind the origin of existing writing systems. Its exploration has led to the development of several visual language systems and is a core part of graphic design assignments, such as icon design.

As is the case with problems from other fields, the visual representation of concepts has also been addressed using computational approaches. In this thesis, we focus on the computational generation of visual symbols to represent concepts, specifically through the use of blending.

We started by studying aspects related to the transformation mechanisms used in the visual blending process, which led to the proposal of a visual blending taxonomy that can be used in the study and production of visual blends. In addition to the study of visual blending, we conceived and implemented several systems: a system for the automatic generation of visual blends using a descriptive approach, with which we conducted an experiment with three concepts (*pig*, *angel* and *cactus*); a visual blending system based on the combination of emoji, which we called *Emojinating*; and a system for the generation of flags, which we called *Moody Flags*. The experimental results obtained through multiple user studies indicate that the systems that we developed are able to represent abstract concepts, which can be useful in ideation activities and for visualisation purposes.

Overall, the purpose of our study is to explore how the representation of concepts can be done through visual blending. We established that visual blending should be grounded on the conceptual level, leading to what we refer to as Visual Conceptual Blending. We delineated a roadmap for the implementation of visual conceptual blending and described resources that can help in such a venture, as is the case of a categorisation of emoji oriented towards visual blending.

RESUMO

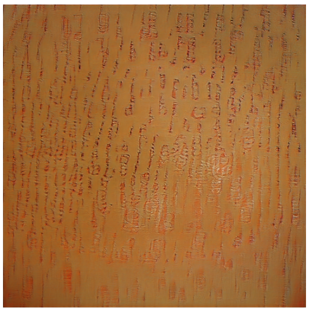
A representação visual de conceitos foi o foco de diversos estudos ao longo da História e é considerada a origem dos sistemas de escrita existentes. A sua exploração levou ao desenvolvimento de vários sistemas de linguagem visual e tem grande importância na área do design gráfico, em tarefas como o design de ícones.

À semelhança de problemas de outras áreas, a representação visual de conceitos também tem sido abordada através de abordagens computacionais. Nesta tese, damos foco à geração computacional de símbolos visuais para representação de conceitos, especificamente através do uso de *blending* ("mistura").[1]

Começámos por estudar aspectos relacionados com os mecanismos de transformação utilizados no processo de mistura visual, o que resultou na proposta de uma taxonomia de mistura visual, possível de ser utilizada para efeitos de estudo e produção de misturas visuais. Além do estudo de mistura visual, concebemos e implementámos vários sistemas: um sistema de geração automática de misturas visuais utilizando uma abordagem descritiva, com o qual realizámos um estudo com três conceitos (*porco*, *anjo* e *cacto*); um sistema de mistura visual baseado na combinação de emojis, o qual chamámos de *Emojinating*; e um sistema para geração de bandeiras, o qual chamámos de *Moody Flags*. Os resultados experimentais obtidos através de múltiplos estudos com utilizadores indicam que os sistemas desenvolvidos são capazes de representar conceitos abstratos e podem ser úteis em atividades de ideação e para fins de visualização.

De modo geral, o objetivo do nosso estudo é explorar como a representação de conceitos pode ser feita por meio de mistura visual. Propomos que a mistura visual deve ter uma base conceptual, promovendo o que chamamos de *Visual Conceptual Blending* ("mistura visio-conceptual"). Delineámos um plano para a implementação de mistura visio-conceptual e descrevemos recursos que podem auxiliar nessa tarefa, como é o caso de uma categorização de emoji direcionada para mistura visual.

1 Traduções alternativas para o termo *blending*: "integração" (Fauconnier e Turner, 1998) e "fusão".



The image shown above is meant to depict a "painting of a student writing a PhD thesis".[2] *I consider it to be a somehow accurate representation, as I would describe the PhD journey as a messy, hard to decode and often "bloody" experience. Reaching the end of my journey was only possible because I was constantly accompanied by people, who helped me in multiple ways and to whom I am forever grateful.*

ACKNOWLEDGMENTS

First and foremost, I would like to thank my two supervisors. I am honoured to have Professor Dr. Penousal Machado as my scientific mentor, who has given me the opportunity and all the conditions to explore my ideas. I cannot describe his supervision better than by using a translation of his own words: thank you for allowing me to freely explore, without imposing restrictions nor paths, but always skilfully guiding me in the desired direction. To Professor Dr. Pedro Martins I owe most of the motivation behind this thesis, being responsible for planting the seed that led to my fascination for Computational Creativity and Blending. I could not have wished for a better advisor – he always knew how to perfectly balance between supervisor and friend, discussing ideas for countless hours, during our bicycle meetings or while enjoying an ice-cream. Thank you for always believing in my work, especially when I had doubts.

The work developed in this thesis stands on the shoulders of giants. The research on *Conceptual Blending* by Professor Dr. Amílcar Cardoso and Professor Dr. Francisco Câmara Pereira was a foundation stone and inspiration to this project. Moreover, I would like to thank Professor Dr. Amílcar Cardoso for his input in decisive stages of my project and also for his company at several conferences.

A word of gratitude to the Computational Design and Visualization Lab. (CDV), where I learned how to be a researcher. CDV Lab. is a place of creative freedom and full of people with whom it is a pleasure to work. Even though I owe a debt of gratitude to all my colleagues at CDV Lab., I would like to express my special appreciation to Ana Rodrigues,

2 I received this image from Professor Dr. Penousal Machado, as a sign of support.

Catarina Maçãs, Evgheni Polisciuc, Filipe Assunção, Sérgio Rebelo and Tiago Martins.

I would also like to thank Philipp Wicke, who unknowingly started this journey with me as a source of inspiration and someone I wanted to meet, and ended up becoming a "scientific partner" and, above all, a close friend.

I would like to also mention Professor Dr. Ernesto Costa, who once gave me a postcard-like piece depicting the words "Everything has a solution" written on a sandy beach right next to the water. Although many interpretations can be made of this image and the statement being a cliché of questionable veracity, it has given me motivation in multiple situations. Thank you for your often jesting yet always wise input.

A special word of gratitude goes to Professor Dr. João Bicker for being an exceptional teacher and for inviting me to become one, which was the first step towards this thesis.

I would like to thank the Department of Informatics Engineering of the University of Coimbra for providing me with the conditions to develop my work. In particular, I would like mention an old friend, Professor Dr. Edmundo Monteiro, who helped with a highly needed final push. I would also like to thank CISUC, in particular Jorge Ávila and Carla Lopes for all their help on multiple occasions.

To my family, who has always shown support and never given up on asking about my work, despite knowing that I would do my best to avoid questions. In particular, I would like to thank my grandparents and father, who have led by example and inspired me to always do my best. Likewise, I thank my mother, Pedro and Henrique for their unconditional support. A special thanks to Ana Maria Xavier, who has welcomed me into her home and provided me with the conditions necessary to finish this thesis.

To all my friends who have been part of my journey (in one way or another), in particular to Filipa Figueiredo, Sofia Baptista, Diogo Moisão, Mariana Moisão, Pedro Rocha and Raquel Moreira.

A brief note of appreciation to Kevin Rodrigues (known as *Worakls*), whose music had a key role in keeping my sanity.

Most importantly, this thesis would have not come to fruition without the support, patience and love from Mariana. Thank you for never giving up on me.

I wish to express my gratitude to the examiners of this thesis (Prof. Lydia B. Chilton, Prof. Pablo Gervás, Prof. Geraint A. Wiggins and Prof. Bernardete Ribeiro) for their insightful remarks and for the enjoyable exchange of ideas. I would like to thank the authors of the multiple images used in this thesis, who have promptly given me the permission to reprint them.

CONTENTS

LIST OF FIGURES

Note: images by other authors were reprinted with permission.

LIST OF TABLES

ACRONYMS

AAC Augmentative and Alternative Communication

AI Artificial Intelligence

API Application Programming Interface

CAD Computer-aided Design

CB Conceptual Blending

CC Computational Creativity

CD Computational Design

CMC Computer-Mediated Communication

CNN Convolutional Neural Network

DNN Deep Neural Network

DVI Dynamic Visual Identity

EA	Evolutionary Algorithm
EC	Evolutionary Computation
EDA	Estimation of Distribution Algorithm
ES	Evolutionary Strategies
GAN	Generative Adversarial Network
IEC	Interactive Evolutionary Computation
ML	Machine Learning
NGSL	New General Service List
PCS	Picture Communication Symbols
PNG	Portable Network Graphics
RMSE	Root Mean Square Error
SOA	Symmetric Object Alignment
SVG	Scalable Vector Graphics
VB	Visual Blending
VCB	Visual Conceptual Blending
ZWJ	Zero-Width-Joiner

1 INTRODUCTION

Images are powerful communication tools, capable of encompassing several levels of meaning that occur simultaneously (Negro, Šorm, and Steen, 2018). The term "image" is used to refer to visual products of intentional human activity, as defined by Carroll (1994). When used for communication purposes, images can be decomposed into visual elements that contribute to their layering of meaning (Van Leeuwen, 2001). While some images are meant to work as windows to what they depict, others are not merely representations of objects but of ideas (McQuarrie, 2008).

As far as the visual representation of concepts is concerned, humans have been doing it since more than two hundred thousand years ago – take, for example, cave paintings from the Neolithic period (Meggs and Purvis, 2012; Santos, Cruz, and Barbosa, 2017). The process of visual representation occurs in multiple ways and can result in images that go from fully pictorial, bearing a resemblance to an existing object, to more abstract, for example using colour to represent a given feature. However, visually representing concepts is a complex task, having an open-ended nature and involving a high degree of subjectivity. Moreover, not every concept is easy to represent and interpretation is always dependent on the knowledge and culture of the observer. Related to this, Tamés (2009) proposes the collaborative development of a database that connects images and concepts, addressing different interpretations and resulting in an evolving visual lexicon that is composed of a dynamic image knowledge base with conceptual annotation. Such resource would be helpful for processes of visual representation of concepts that are common, for example, in Graphic Design assignments, in which designers seek to produce symbols that are easy to perceive (e.g. signage systems, user interface icons, etc.).

Of the different types of image-making processes, one is commonly referred to as *Visual Blending* (*VB*) and consists in combining two or more existing images to produce a new one. This type of process is often used in visual language systems to obtain new symbols from existing ones, allowing the representation of new concepts (Horton, 1994). Visual Blending can also be related to the cognitive operation known as *Conceptual Blending* (*CB*), in which two input mental spaces are brought together, resulting in a new blended space (Fauconnier and Turner, 2002). This operation is key in understanding the emergence of meaning. Visual blending can be seen as being based on a conceptual blend, which is expressed at the representation level. In this thesis, we argue in favor of a stronger connection between the two types of blending.

The two types can be used in combination – leading to what we refer to as *Visual Conceptual Blending* (*VCB*) – to enable the creation of visual blends with a solid conceptual grounding, which can be explored in the visual representation of concepts. We consider that the use of perceptual features (e.g. *colour*) is fundamental to exploiting semiotic aspects that are relevant in the visual representation of concepts. For example, the advantages of using visual characteristics related to the data being represented are well documented (Jahanian, 2016a,b), especially in the field of *Information Visualisation*, e.g. the use of semantically-resonant colours has been shown to improve chart reading speed (Lin et al., 2013).

Similarly to other problems, the visual representation of concepts has also been explored using computational approaches. Existing explorations can be framed in the context of several research fields. First, from the perspective of *Computational Design* (*CD*), strategies can be used with the goal of improving a given aspect in the process of producing visual representations. The capabilities of computational approaches in image production have been a topic of discussion. Some authors highlight the advantages of using computers in art and design (e.g. Whale, 2002), while others point out their limitations – e.g. Calude and Lewis (2012) discuss the issues involving the implementation of a universal image generator. Regardless of existing limitations, we believe that, in the case of concepts, automation techniques may be employed in the exploration of the conceptual space in combination with processes of analogy-making (French, 2002) and semiotic analysis, which can result in the generation of visual representations. In this sense and as Whale (2002) states, computers can be used to extend our capabilities. Cunha et al. (2015) argue that computational systems for the visual representation of concepts can be used in ideation activities to foster the creativity of the user. The same notion is mentioned by other authors, such as Karimi et al. (2018b) and Zhao et al. (2020).

Second, from the perspective of *Computational Creativity* (CC), artificial agents that exhibit creative behaviour can be used in the production of visual representations. Such agents may help in the exploration of the search space and identification of good solutions based on a given sense of quality – e.g. using Evolutionary Computation (EC) techniques. Creative systems may present the user with solutions that are surprising, which can lead the user to explore other ideas. Moving past the existing discussion on whether computer systems can be authors (Audry and Ippolito, 2019; Hertzmann, 2018), our position is that Artificial Intelligence (AI) may provide us with new tools to express ourselves. The type of relationship between user and system is also worth exploring, which leads us to the third perspective: *Co-creativity*. Collaborative ideation is considered to aid in the generation of creative solutions, by exposing people to ideas different from their own (Chan et al., 2017; Siangliulue et al., 2015). In *Computational Co-*

creativity, the interaction between human and artificial agents is explored to promote the creative process. In specific, the autonomy of the artificial agent can be explored with the goal of developing systems that use mixed-initiative interaction (Yannakakis, Liapis, and Alexopoulos, 2014), in which both the computer and the human user have a proactive contribution in the creative process. This collaborative relationship can foster the creativity of the user, as addressed by Liapis et al. (2016). Our motivation is in line with the ideas expressed by Veale and Cardoso (2019, p. 2), who see the future of intelligent computers as a transformation from passive tools into active co-creators.

With this thesis, we intend to explore CD techniques and further extend the application of CC in the fields of art and design, in particular related to the visual representation of concepts.

1.1 RESEARCH GOALS

The title of this thesis makes use of term "concept-representative symbols," which can be seen as a pleonasm (a symbol represents a concept by definition), to highlight our main goal: explore how computational approaches can be used for the visual representation of concepts. As such, our research hypothesis is that *computational approaches can be employed to produce visual representations of concepts, which can be useful for fostering creativity in ideation activities and for facilitating comprehension in visualisation contexts*. Based on this research hypothesis, the following research questions arise:

A. *Can Computational Approaches, in particular those based on Visual Blending, be used for the visual representation of concepts?*
Multiple visual aspects have been explored in computational approaches to encode meaning. We are interested in studying how this encoding can be done with the goal of representing specific concepts. Our focus is on a particular process of producing images: visual blending. We question which aspects need to be considered in visual blending and how a bridge can be built between the visual and the conceptual levels, possibly leading to visual conceptual blending.

B. *How can the user be integrated and allowed to express their preferences?*
The production of visual representations of concepts is closely related to the preferences, goals and general context of the agent who is responsible for the representation. Therefore, in a context of computational design, the preferences and goals of the user should be taken into account, allowing the user to influence the system and have some control over the final output. We question how this control from the user can be implemented and which degree of user-dependency should be encoded into the system (from fully autonomous to co-creative).

c. *How are the generated symbols perceived by users?*
One of the main concerns in the production of visual representation of concepts involves their perception by a viewer. This topic is related to how symbols are interpreted. We question how users perceive the generated symbols and whether the meaning encoded by the system can be extracted.

To address the aforementioned research questions, we pursue the following objectives:

1. Study the state of the art in the visual representation of concepts, especially concerning computational approaches;
2. Identify application domains to explore the visual representation of concepts;
3. Conceive and implement approaches for the generation of concept-representative symbols;
4. Explore techniques for the exploration of conceptual search spaces;
5. Study visual blending techniques for concept representation;
6. Explore the use of semiotic-related characteristics for concept representation;
7. Employ techniques for the automatic assessment of the quality of generated symbols;
8. Explore methods to allow the user to express preferences;
9. Assess the impact of the developed systems in ideation and visualisation;
10. Assess the quality of the developed approaches in regards to visual conceptual blending.

In order to achieve these objectives, we explore different approaches, which help us to iteratively tackle the problem at hand. The end goal is the development of systems for the computational generation of visual representations of concepts. Our focus resides in the use of blending techniques, from the conceptual level (conceptual blending) to the visual one (visual blending). We also employ methods to allow the user to take advantage of the system, for example by using Interactive Evolutionary Computation (IEC) techniques or establishing a cooperative relationship between human and computer, through co-creative approaches. With this, we aim for the developed systems to be useful in design ideation and in visualisation activities. To test the developed systems, we conduct several studies, some of which involving users and following mixed-methods approaches, which combine quantitive and qualitative methods (Venkatesh, Brown, and Bala, 2013).

1.2 CONTRIBUTIONS

The work developed in the context of this thesis resulted in contributions to several fields. In this section, we outline the main ones.

LITERATURE REVIEW: We conducted a review of the state of the art addressing topics related to the visual representation of concepts (Chapter 2) – *Visual Perception* (Section 2.1), *Concepts* (Section 2.2), *Semiotics* (Section 2.3) and *Visual Grammar* (Section 2.4) – and computational approaches to it (Chapter 3), *Conceptual Blending* (Section 4.1) and *Visual Blending* (Section 4.2). We also analysed topics related to our practical research, for example, *Emoji* (Chapter 7), *Glyphs* (Sections 13.1 and 13.2), *Flags* (Sections 14.1 and 14.2) and *Image Schemas* (Section 16.1). The base of our investigation was described in an initial publication (Cunha et al., 2015).

VISUAL BLENDING TAXONOMY: We analysed existing taxonomies related to visual blending and proposed a novel one, based on the transformation mechanisms used to produce visual blends. The proposed taxonomy is useful both for the analysis of visual blends and the production of new ones. This contribution is described in Chapter 5.

STUDY ON VISUAL BLENDING: We conducted an analysis of two image datasets that include visual blends (*VisMet* and *Emoji Kitchen*), focusing on the different types of transformation used. In this analysis, we used the developed taxonomy to identify the most common transformations in the visual blends of the two datasets. This contribution is described in Chapter 5.

VISUAL BLENDING SYSTEM BASED ON DESCRIPTIVE APPROACH: We developed a system for automatic generation of visual blends using a descriptive approach, which has a hybrid blending process, starting at the conceptual level and ending at the visual one. We conducted experiments using three concepts (*pig*, *angel* and *cactus*) and tested the system with a user study focused on perception. This contribution is described in Chapter 6 and resulted in one publication (Cunha et al., 2017).

EMOJINATING: We iteratively developed a visual blending system that we called *Emojinating*, which uses the blending of emoji to visually represent concepts. Three main versions of the system were developed: deterministic (Chapter 8), evolutionary (Chapter 9) and co-creative (Chapter 11). We conducted user studies with all versions. A special reference should be made to two of the studies: a study that compared the performance of the system in the representation of single-word concepts with double-word concepts (Chapter 10); and a study focused on usefulness and perception (Chapter 12). This contribution resulted in several publications (Cunha et al., 2019a; Cunha et al., 2020b; Cunha, Martins, and Machado, 2018a,b, 2020a; Cunha et al., 2019b).

STRATEGY FOR RETRIEVAL OF DATA-RELATED GLYPHS: Based on the *Emojinating*'s engine, we proposed a strategy to be used in the context of Information Visualisation to retrieve glyphs related to the data thematic. This strategy has high application potential for a visualisation tool, allowing it to have different types of glyphs suitable to different thematics. This contribution is described in Chapter 13 and resulted in one publication (Cunha et al., 2018).

SYSTEM FOR FLAG GENERATION: We developed the system *Moody Flags*, which generates flags based on trending topics of countries, retrieved from real-time news. The development of the system required the construction of a dataset of semantic and visual flag data. We also conducted a user study to assess the perception of generated flags by users. This contribution is described in Chapter 14 and resulted in two publications (Cunha, Martins, and Machado, 2020b; Cunha et al., 2020c).

VISUAL CONCEPTUAL BLENDING ROADMAP: We proposed a roadmap for the implementation of visual conceptual blending, highlighting useful resources. This contribution is described in Chapter 15 and resulted in two publications (Cunha, Martins, and Machado, 2018c; Cunha, Martins, and Machado, 2020d).

EMOJI CATEGORISATION FOR VISUAL BLENDING: We proposed an emoji categorisation oriented towards visual blending, which can be used in the implementation of a system for visual conceptual blending. We validated the categorisation with a user study and used it to analyse the *Emoji Kitchen* dataset with the goal of identifying transformational patterns in emoji categories. This contribution is described in Chapter 17.

In addition to the international peer-reviewed publications identified next to each contribution, we also published material related to the general topics of the thesis (Cunha and Cardoso, 2019; Cunha, Martins, and Machado, 2020c) and contributed to several supporting publications (Cruz, Hardman, and Cunha, 2018; Lopes, Cunha, and Martins, 2020; Wicke and Cunha, 2020).

1.3 DOCUMENT OUTLINE

In this chapter, we have introduced the motivation for the research described in this thesis, as well as its goals and contributions. The remainder of the document is divided into six parts, which we describe in the following paragraphs.

In Part I, we describe the state of the art on the visual representation of concepts. In Chapter 2, we introduce the reader to *Visual Perception*, *Concepts*, *Semiotics* and *Visual Grammar*. In Chapter 3, we start by introducing the areas of *Computational Design*, *Computational Creativity* and *Computational Co-creativity*. Then, we review existing computational approaches that address the visual representation of concepts.

In Part II, we present Blending as one of the main focuses of the thesis. In Chapter 4, we provide an overview of two sides of blending: conceptual (concept combination and conceptual blending) and visual (visual metaphor, visual structure, perceptual features and transformational perspective). In Chapter 5, we present an analysis of visual blends from two different image datasets (*VisMet* and *Emoji Kitchen*), in which we aim to identify usual transformations. In Chapter 6, we describe the development of a system for the automatic generation of visual blends using a descriptive approach.

In Part III, we explore the visual representation of concepts through visual blending of emoji. In Chapter 7, we introduce the reader to emoji, pointing out their potential for visual representation of concepts. Then, we describe the iterative development of a visual blending system that we called *Emojinating* – Chapter 8 presents the deterministic version, Chapter 9 presents the evolutionary version and Chapter 11 presents the co-creative version. In Chapter 10, we describe a study that assessed the performance of *Emojinating*, comparing its use with single-word concepts and double-word concepts. In Chapter 12, we make an overall analysis of *Emojinating* and we describe a user study focused on usefulness and perception.

In Part IV, we focus on other domains in which visual representation of concepts has high application potential. In Chapter 13, we explore the use of *Emojinating*'s engine in a context of Visualisation. In Chapter 14, we describe the implementation of *Moody Flags*, a system that generates flags based on trending topics of countries, retrieved from real-time news.

In Part V, we address open questions related to the visual representation of concepts. In Chapter 15, we propose a roadmap for the implementation of visual conceptual blending systems. In Chapter 16, we describe how affordance-related features can be considered in systems for the visual representation of concepts by using image schemas. In Chapter 17, we present the development of a categorisation of emoji oriented towards visual blending, which can be used in visual conceptual blending systems.

Finally, Part VI comprises Chapter 18, in which we summarise the work developed in the scope of this thesis. We start by revisiting the research questions, highlighting the main contributions. Then, we conclude with an overview of the thesis parts.

Throughout the document, we often use the abbreviation "blend" to refer to visual blend and "representation" to refer to "visual representation". The webpages cited in this document were available at the time of writing (2021-2022), unless explicitly stated otherwise.

Part I

STATE OF THE ART

2 VISUAL REPRESENTATION OF CONCEPTS

Humans have been visually representing ideas since more than two hundred thousand years ago. Take, for example, cave paintings and petroglyphs (Fig. 2.1). These representations vary from being fully pictorial and depicting concrete things, e.g. pictograms, to more abstract and depicting ideas, e.g. ideographs. In any case, they can be considered visual representations of concepts.

In order to make clear what is meant by "visual representation," we borrow from the definition of "graphic representation" by Engelhardt (2002), which in our perspective can be viewed as similar.

> *A graphic representation is a visible artefact on a more or less flat surface, that was created in order to express information.*
> (Engelhardt, 2002, p. 2)

Engelhardt (2002) identifies as a requirement of graphic representation that there is a clear intention to express information, no matter the medium in which it occurs. This requirement includes, for example, cave paintings but excludes what is referred to as "self-occurring 'natural signs"', such as footprints in the sand. Moreover, Engelhardt (2002) distances "graphic representation" from artefacts with "the intention to amuse, delight, persuade, invigorate, provoke or otherwise stimulate", dismissing images in advertising and art. A note should be made on this aspect, as we consider that such images can be viewed as visual representations, as long as they were produced to encode and express information, no matter their degree of abstraction, the difficulty of meaning decoding or even whether their main goal is to amuse rather than to inform.

Figure 2.1: Details of the painted dolmen of Antelas, example of megalithic art from the Neolithic period located in Oliveira de Frades, Portugal. Image produced by Fernando Barbosa, published by Santos, Cruz, and Barbosa (2017).

Moreover, an important distinction is made by Bertin (2011) between two types of graphic representation: graphics and pictography. The former concerns the representation of predefined sets, represented by a data table, and is subdivided by Bertin (2011) into diagrams, networks and maps. The latter is described as being related to the definition of a set (or concept) in the reader's mind (Bertin, 2011), which is linked to the design of symbols by Engelhardt (2002). In the context of this thesis, we focus on what is referred to as pictography.

To put it briefly, we employ the term "visual representation" instead of "graphic representation", as we consider it to be more appropriate due to several reasons: (i) "graphic" is often related to the use of elements such as *line*, *shape*, *type*, etc., while "visual" is more general and easily connected with other visual media that can also be used in the representation of concepts, e.g. photography; (ii) "visual" is less likely to be mistaken with "graphics"; (iii) and is closer to the perceptual channel referred to as "visual".

In this chapter, we introduce topics that are key to the understanding of visual representation of concepts. We start by describing aspects of *visual perception*, such as *visual properties* (e.g. *colour*) and principles (e.g. *Gestalt Principles*). Then, we shift our attention to concepts and describe what they are, how they are considered to be represented cognitively and which conceptual aspects should be taken into account when visually representing them. Afterwards, we make a brief introduction to the perspective of *Semiotics* and describe how visual aspects can be related to meaning. In the last part of the chapter, we show how all these concepts come together by introducing the notion of *visual grammar* and presenting examples of *visual language systems*.

2.1 VISUAL PERCEPTION

When we look at a graphic representation, a mental construction is built through mechanisms of visual perception. Visual perception can be described as the process by which visual information is collected and processed, involving a complex interaction among various stimuli. In this process, a search for patterns is conducted with the goal of identifying meaningful structures (Pettersson, 2011). A similar notion is proposed by Marr (1982), who considers perception as the transformation of a pattern of light on the retina into awareness of the visual world (Bruce, Green, and Georgeson, 2003, p. 80).

Zimbardo and Gerrig (2002) state that the process of perception is best understood if divided into three stages: sensation; perceptual organisation; and identification/recognition of objects. Similarly, Ware (2012) describes three stages of perceptual processing. In the first stage, *low-level properties* (e.g. *orientation*, *colour*, *texture*, etc.) are extracted from the environment and processed unconsciously by neurons. This stage is related to the ideas from Marr (1982), who describes low-level vision as an autonomous process that results in a symbolic representation, useful for higher-level processes – introducing the term *primal sketch*. In the second stage, the visual scene is divided into regions and simple patterns. These patterns can be found, for example, in elements or regions with the same colour. The third and last stage consists in focusing on a reduced number of objects, which are held in the visual working memory. A few aspects can be highlighted from this three-stage model, from visual properties to principles of *Gestalt* perception. We will devote our attention to them in the following sections.

2.1.1 *Visual Properties*

The first aspect is that there are certain properties that are retrieved from the environment in the process of visual perception.

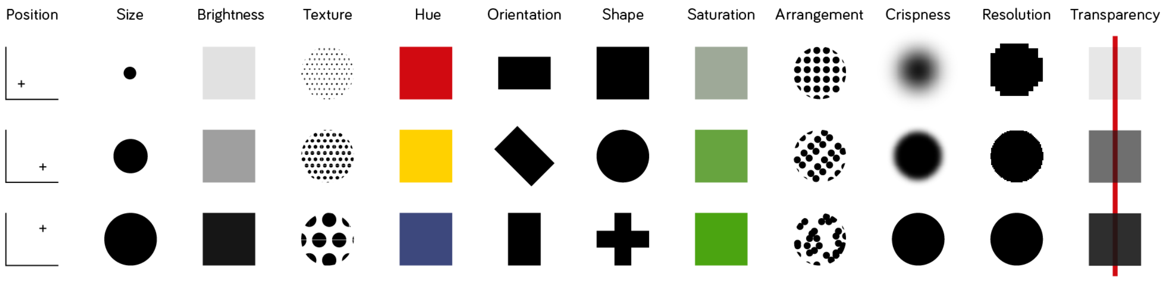


Figure 2.2: Visual variables. Adapted from: (Polisciuc, 2021).

Related to this topic and coming from a perspective of data visualisation, Bertin (2011) proposed an approach to structure the visual representation of data. In their proposal, information is encoded by using visual elements (*marks*) and varying their *visual properties*. Bertin (2011) identifies seven *visual variables*: *position*, *size*, *texture*, *value*, *colour*, *orientation* and *shape*.[1]

The terminology used to refer to *visual properties* (also referred to as *visual variables* and *visual attributes*) differs from author to author. Bertin's "value" refers to what is commonly known as *brightness* (light versus dark) and "colour" can be divided into *hue* and *saturation* (Engelhardt, 2002). In addition, Bertin's list of visual features was later extended by other authors. For example the addition of *arrangement*, *transparency of fill* and *crispness* (*fuzziness*) *of edges* (MacEachren, 2001). Engelhardt (2002) mentions that "texture" can be further subdivided into *size* of texture elements (similar to Bertin's definition), *shape* of texture elements and *orientation* of texture elements. A summary of visual properties can be seen in Fig. 2.2.

Engelhardt (2002) divides visual properties into two groups: *area-fill* (*brightness*, *hue*, *saturation* and *texture*) and *spatial* (*shape*, *size* and all the remaining ones). This division is based on the idea that a change on a spatial attribute alters the anchoring of the object or part of it, while a change on an area-fill attribute does not.

These visual variables are introduced as properties that can be used to encode meaning in the context of data visualisation. Nonetheless, they are aligned with the properties mentioned by Ware (2012).

2.1.2 *Perceptual Principles*

Beyond the individual visual properties, the second stage proposed by Ware (2012) is related to pattern identification. The process of pattern identification is considered to be conducted according to certain principles, for example how we distinguish between *figure* (the most salient elements) and *ground* (the background elements). These principles,

1 It is important to notice that the work by Bertin (2011) goes beyond the identification of visual variables, describing, for example, how organisation levels have a direct impact on visualisation.

known as *Gestalt Principles*, were initially addressed by a group of German psychologists (Köhler, 1929; Wertheimer, 1938) and are based on the idea that perception does not simply rely on the analysis of individual properties but has instead a holistic nature – the whole is different from the sum of its parts.

The principles are considered to be the base for how we perceive wholes from perceptually individual elements, in an unconscious manner. For example, the *principle of proximity* shows how elements that are positioned closer together can be perceived as a group. Another example is the *principle of similarity*, according to which elements with similar visual features (e.g. *size*, *shape* or *colour*) are perceived as belonging to the same group. Other principles exist – e.g. the *principle of closure*, the *principle of continuity*, the *principle of smallness* – we refer the reader to Pettersson (2011) for further details.

Arnheim (1974a) writes about how these principles are applied in art and provides insights on other aspects of perception, such as *depth representation* (e.g. creating depth using gradients).

On a higher level, a composition can be perceived as having certain aesthetic qualities. A survey by Stebbing (2004) identifies four *organisational principles*: *contrast*, *rhythm*, *balance* and *proportion*. The work by Graves (1951) is relevant on this topic, describing how these principles can be explored through the use of *visual properties*, giving examples from simple shape interaction to more complex artworks.

Even though these principles are important when studying perception, for the purpose of this thesis, the focus is more on how specific visual properties can be assigned with meaning and used to represent concepts. With this in mind, we now shift our attention to concepts.

2.2 CONCEPTS

The topic of this thesis is centred on how concepts can be visually represented. As such, it is important to first explain what we mean when we use the term "concept".

A *concept* can be defined as a dynamic abstraction that refers to ideas, objects or actions (Pereira, 2004). Although concepts may be considered somehow static, they are often subject to change. This change may be caused by time (e.g. technology may bring changes to our understanding of a given concept) or vary according to context (e.g. the concept of "near") and experience/culture (e.g. in some cultures "cow" is considered a "common" animal, in others it is viewed as "sacred").

A concept can be equated with the set of things that it refers to, for example, the concept which we refer to when we use the word "flower" is equal to the set of things that we consider flowers. This is the base for referring to concepts as "cognitive categories". As such, the term "category" is used to refer to a set of things that are equivalent at some level (Rosch et al., 1976). These categories are seen as the result of a mental process of categorisation (Ungerer and Schmid, 2006).

When we refer to these *cognitive categories* (or *concepts*) we use words. However, it is important to bear in mind that they are not equivalent (Ungerer and Schmid, 2006). For example, the Portuguese word "cadeira" and the English word "chair" are used to refer to the same object that a person can sit on. However, a given word can be used to denote several categories, e.g. the English word "chair" can also mean the president of a meeting, a meaning that is not covered by the Portuguese word. As such, words should not be seen as concepts but only as symbols that refer to them. Moreover, these symbols (words) are used not only to refer to categories of things that physically exist (e.g. "chairs") but also to more abstract entities (e.g. actions, emotions, spatial relationships, etc.) and even categories that are made up on the fly on a given subject under consideration (e.g. "things to take from one's home during a fire") (Barsalou, 1983).

One of the main topics studied about concepts is how they are cognitively represented. Although several views exist, e.g. *Exemplar View* (Medin and Schaffer, 1978) or *Theory View* (Murphy and Medin, 1985), we will focus on the *Prototype Theory* (Rosch, 1975), which provides insights of particular relevance for this thesis.

2.2.1 *Categories and Prototypes*

In the classical view, categories are seen as definable by common properties, have clear boundaries and binary membership (Lakoff, 1990). The classical view is challenged on several fronts but it is the work by Rosch (1975) that provides a more solid alternative, commonly known as *Prototype Theory*.

According to the classical view, all members of a category have equal status, sharing the properties that define the category. On the other hand, Rosch and Mervis (1975) demonstrate that categories have *prototypes*, which are considered as best examples. According to Rosch et al. (1976), thought is organised in terms of prototypes and basic-level structures. First, for a given category, some members can be seen as "better examples" than others. Second, categories are considered to have degrees of membership and no clear boundaries. This view is in line with studies that show that categories are graded. An example is the work by Labov (1973), who explored the semantic boundaries of cup-like containers by presenting subjects with line drawings that varied in width and height (Fig. 2.3). The results of their experiments show that the boundaries between neighbouring categories (*cup*, *mug*, *bowl* and *vase*) are fuzzy and context-dependent (results differ when the containers are considered in a different context, e.g. a *food* context in which they were filled with mashed potatoes).

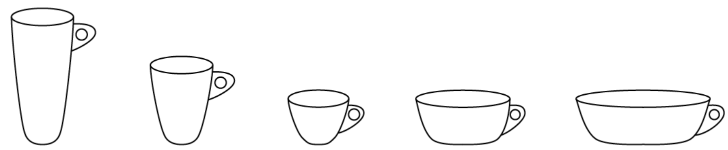


Figure 2.3: Drawings of cup-like objects. Adapted from: (Ungerer and Schmid, 2006).

The *Prototype Theory* does not clarify how to draw boundaries between categories but only how one can identify the most typical members, which have the average features of the category. A list of features is used to define the prototype of each concept. These prototypical features are used in our process of understanding the world. For example, the prototype of *fish* would include that fish have *gills* and *scales*, that they *swim* and are *cold blooded*. This prototype is used to assess if a given instance is an example of *fish*. Considering the prototypical features of *fish*, we can reach the conclusion that a *whale* would not be considered an example as, despite swimming, it is warm-blooded, has lungs instead of gills and does not have scales. In addition, the members of a category do not all share the same features, e.g. compare a penguin with an ostrich. Instead, they are linked by family resemblances (Rosch and Mervis, 1975) in the sense that each member has some similarities to other members, which provides coherence to the category. We have used the term "prototypical feature" to refer to typical characteristics of a given category. This term is similar to what Rosch et al. (1976) refers to as "cue validity" – an assessment of the frequency with which a given cue is associated with a category and also with others, in which cue validity for a given category is highest when the cue has high frequency in the category and low in others. Similar ideas are explored by other authors. For example, Costello and Keane (2000) use the term "diagnostic predicate" in reference to features that best identify instances of a given concept and differentiate it from others – the example of *cacti* is given, identifying *prickly* as more diagnostic than *green*.

A given object is seen as belonging to a category if it bears sufficient resemblance to the category's prototype. Although a *penguin* is very different from an *ostrich*, most people would easily classify them both as *birds*. On this topic, Lakoff (1990) summarises that some categories have clear boundaries (e.g. *birds*) but within the boundaries there is a gradation from a prototypical member, while others have fuzzy boundaries and inherent degrees of membership (e.g. *red*, *tall man*).

Another aspect of the *Prototype* view is that categories are considered to follow a taxonomic hierarchy (Rosch and Mervis, 1975), in which they are related to another by a means of inclusion – the more inclusive a category is, the higher is its level of abstraction. As such, different levels of categories exist, from more general (e.g. *animal* and *furniture*) to more specific (e.g. *retriever* and *rocking chair*). In the middle,

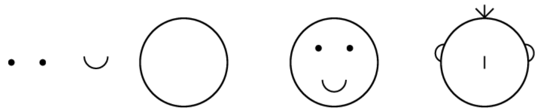


Figure 2.4: Reduction of face

a *basic level* is considered to exist (e.g. *dog* and *chair*). This inclusion-structured hierarchy is based on *type-of* relationships, e.g. a *retriever* is a type of *dog*. This basic level is considered the highest and most inclusive level at which category members have a significant number of common attributes, have similarly perceived shapes and possess similar motor programs, being interacted with using similar motor actions (Rosch et al., 1976). The basic level is also considered the first to be learned by children. At this level, categories maximise the perceived similarity among members and minimise perceived similarities across contrasting categories (Lakoff, 1990).

2.2.2 *Reduction to Essential*

In the cup experiment (Labov, 1973), the width/height ratio was shown to have an important role in how the object was recognised. This demonstrates the importance of the features of an object in how it is perceived.

By focusing on object composition, it is possible to further investigate our perception. Ungerer and Schmid (2006) state that, although objects are perceived as wholes, their parts have an important role in establishing a prototypical exemplar for a given concept. The *Recognition by components Theory* (Biederman, 1987) is also related to this idea, proposing that objects can be seen as a combination of a limited set of simple geometric shapes (e.g. cones and cubes), referred to as *geons*. For example, the prototypical representation of a dog can be based on a combination of an ellipsoidal head, a cylindrical torso, four cone-shaped legs and an expanding handle for a tail (Hedblom and Kutz, 2015).

Ungerer and Schmid (2006) mention that for some objects, a prototypical representation can be based on a reduction to relevant parts. The example of a bungalow is given, which can be represented by simple shapes depicting a one-storey building with only walls, roof, windows and a door. Overall, one can say that there is an inclination towards simplicity, which can be observed in the use of schematic images in dictionaries, in the *principle of parsimony* in science, which states that a theory is preferred when it is simpler (Arnheim, 1974b), and even in the reduced periods in which *Realism* overlaps *Abstractionism* throughout art history (Machado, 2007, p. 218).

Nevertheless, the visual representation based on simplification and essential elements is far from being trivial. First, one must identify which elements should be considered in the representation. For example, when representing a face, one has to decide which elements to include. However, the elements on their own are not themselves sufficient but it is their combination and positioning that triggers an interpretation (Hassan, 2015), e.g. the dots alone are not enough to represent a face (see Fig. 2.4). Moreover, the elements do not all have the same importance: a mouth is considered to be diagnostic of face (i.e. if there is a face, there is a mouth), whereas hair for example is not (Schilperoord, 2018). For this reason, a representation with eyes and mouth is more characteristic of face than a representation with hair and ears (see Fig. 2.4). In addition to a reduction to parts, one needs to take into consideration other aspects – e.g. parts not only contribute to the overall shape of the object but are also related to its *function*. On this topic, Lakoff (1990) points out that we interact with objects using their parts, which are key in defining the *motor programs* that one can use – cognitive structures developed when learning movement (Raiola, Tafuri, and Gomez Paloma, 2014). For example, the relevant parts of a chair are its legs, seat and back. Therefore, the image of a prototypical chair is likely to rely on the presence of these parts. However, for the prototypical representation of a chair, it is also important that these parts have a proportion that is considered optimal in terms of their function (Ungerer and Schmid, 2006). Figure 2.5 shows two chairs with different proportions – the top one is closer to the average look of a chair, which makes it likely to be seen as more prototypical, whereas the one on the bottom has proportions that make it similar to a *chaise longue*.

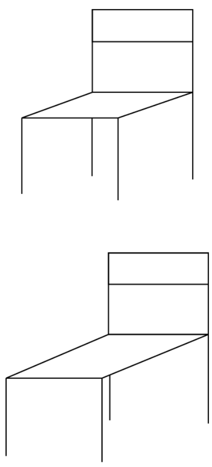


Figure 2.5: Chairs

Therefore, function is a key aspect in the way objects are perceived. We refer again to the cup experiment by Labov (1973) in which different results were obtained depending on the object's supposed contents. Related to this idea, Mandler (2000) identifies two types of cognitive processes that occur in categorisation: perceptual and conceptual. While the former concerns the perceptual similarity among objects, the latter is focused on what the objects do or what they are used for. Although the cup-like objects shown in Fig. 2.3 depict what can be perceived as cups, they do not cover the full extent of the concept. For example, for *cup of coffee*, the perceptual side is important but it fails to capture the essence of the concept, which can only be achieved if one considers the purpose of containing *coffee*. The *Theory of Affordances* by Gibson (1977) addresses this issue by highlighting the importance of usages and purposes in our perception of objects. Ideally, a visual representation should be able to capture these two levels.

Going back to the examples of *bungalow* and *chair*, by looking at their simplified visual representations and 2.5), an observer would likely be able to perceive the concepts behind them, even considering that they do not depict affordances. This could be seen as support to the idea that a method based on a reduction to basic shapes can be used to visually represent concepts. However, it is easy to reach the conclusion that a generalisation of this method is not feasible. While some concepts are fully grounded on visual aspects, e.g. *red*, and others strongly rely on their visual appearance, e.g. *dog*, there are also others that have a more complex nature. For example, although the diagnostic aspects of *measles* are visual (a red rash), they do not suffice for a full understanding of the concept, which is supported by the definition of *virus*. On the other hand, some concepts are even complex to understand and define, let alone visually represent, e.g. *game*. Even focusing on concrete concepts, the problem persists: some are too complex to reduce (e.g. an *English pub*) or have key features that are not merely visual (e.g. *teddy bear* is commonly linked to *softness*). All in all, not every concept has a direct translation into a visual representation.

2.2.3 *Conceptual Connections*

As already shown, some concepts can be hard to visually represent. This is often the case with superordinate categories (e.g. fruit or animal), whose members are not linked based on appearance (e.g. a common shape). In such cases, a representation may rely on properties from related categories. For example, when asked to represent "fruit", it is very likely that one would resort to drawing a specific fruit, e.g. a banana. Similarly, from a teddy bear pictogram in a store, one can easily infer that it refers to the toys section. This mechanism consists in borrowing properties from more specific categories and is referred to as parasitic categorisation (Ungerer and Schmid, 2006).

The inverse may also happen in the case of specific concepts, resorting to attributes of superordinate categories, even though these are shared by other categories. For example, to understand the concept *white-flowering Japanese dandelion*, one would likely resort to their knowledge on *dandelion* and, upon being able to categorise it as a *flower*, one could also draw information from *flower*. Subordinate categories are based on the highlighting of specific attributes (e.g. the *white-flowering*). As one goes up to a higher level (e.g. *flower*), only general attributes are seen as salient enough (e.g. has a stem, petals and leaves).

These two mechanisms explore *type-of* relations, e.g. a *dandelion* is a type of *flower*, which are the base for the already mentioned hierarchy of categories. However, this is just one of the types of hierarchies that can be established.

As another example, consider the concept *house*. The concept *house* can be said to be related to *bedroom*, *kitchen* and *bathroom*. Moreover, one can establish a relation of *part-whole* among them – *bedroom*, *kitchen* and *bathroom* can be considered as parts of a *house*. In this case, a *basic level* category (*house*) is considered to be used as *superordinate*. However, its *function* is not to highlight certain attributes of its members (as is the case of *type-of*) but has an "assembling" role, in which the members compose the *superordinate* (e.g. house can be composed of a bedroom, a kitchen, etc).

Ungerer and Schmid (2006) addresses the topic of conceptual hierarchies and refers to the *type-of* and *part-whole* relations as two ways of establishing hierarchies. In the case of *type-of*, hierarchies are established based on class inclusion and salient attributes. In *part-whole*, the hierarchies are based on a relationship of *spatial contiguity*, *connectivity* and *continuity*. These connections among categories are useful and may be explored in concept visual representation, as we have seen with the case of *toy* represented using an image of a teddy bear.

The *part-whole* relation is often behind a type of conceptual connection referred to as *metonym*. In metonym a given concept stands for another one, based on correlations such as *part-whole*, *inside-outside* or *cause-result* (Ungerer and Schmid, 2006).

In addition to the already mentioned types of relations, it is also possible to establish others. Consider, for example, the concept *eye*. It is easy to connect it with *face*, on a basis of a *part-whole* relation, or with *organ*, on a basis of a *type-of* relation. However, one can also connect it with the concept *camera*, based on comparison. This type of relation is grounded on *analogy*, which consists in a mapping of knowledge between two domains, supported on relations between objects (Gentner and Jeziorski, 1993). Analogies can lead to metaphors, such as "camera as eye", which is central in cinema (Quendler, 2016).

Although initially approached as purely linguistic, metaphors are currently considered as "cognitive instruments" that allow us to think of a given concept using another one (Black, 1962, p. 237). In metaphor, a mapping occurs between a source model, e.g. eye, and a target model, e.g. camera, leading to camera being seen as an eye.[2]

A mapping process regulates the eligibility of correspondences between the two domains that lead to the metaphor. Ungerer and Schmid (2006, p. 121) distinguish three major components of mapping scopes: image schemas, which are grounded in our bodily experiences, e.g. the UP-DOWN schema can be related to quantity (Vernillo, 2018), more is up and less is down; basic correlations (e.g. *cause-effect*, *purpose-goal*) and culture-dependent evaluations (e.g. a common evaluation is to see money as something valuable).

Despite both being seen as cognitive instruments, *metaphor* is considered more complex than *metonym*. For example, in metonym the range of *source* and *target* concepts is normally seen as restricted to concrete concepts (Ungerer and Schmid, 2006). Interestingly, Bolognesi and Vernillo (2019) study how metonymy can be used in the visual representation of abstract concepts through concrete ones, proposing the principle of *Abstraction by Metonymy*. In any case, they are both useful

2 As is the case with other conceptual mappings, the relation between camera and eye can lead to two metaphors: "camera as eye" and "eye as camera". The visual representation of these metaphors may not be the same.

in facilitating understanding and also representation. Take, for example, the concept *argument*, which may not be easy to visually represent. However, it is possible to conceptualise it in different manners, for example as *clash* or *quarrel*. In this way, for its representation, one could resort to a similar concept.

2.3 SEMIOTICS

Having presented components of visual perception in Section 2.1, and defined concepts and aspects that should be considered when addressing their visual representation in Section 2.2, we now bring the two together by addressing how meaning can be encoded visually. The field devoted to studying such matters is called *Semiotics*.

> **Semiotics:**
> *the study of signs and symbols and of their meaning and use.*
> (as defined by the *Oxford Learner's Dictionaries*)[3]

As the definition states, *Semiotics* deals with how signs convey meaning, connecting aspects of perception with the conceptual level. In addition, signs can be communicated through multiple senses, e.g. *visual*, *auditory*, etc. In the case of this thesis, we focus on visual signs. Moreover, of all visual signs, we are interested in the ones that are within the scope of visual representation, i.e. artefacts created "on a more or less flat surface" with the intention of expressing information (as defined at the beginning of the chapter). In this section, we will introduce *Semiotics* and present aspects that are relevant for the visual representation of concepts.

2.3.1 *Aspects of Semiotics*

At the beginning of this chapter, we described visual properties that are available when producing representations. However, these properties are not merely visual but go hand in hand with meaning. As such, it is important to highlight how the meaning of a given representation can be changed by altering some of its visual characteristics.

Although we presented multiple visual properties in Section 2.1.1, some of them are especially relevant when conveying meaning, for example *shape*, *colour* and *position*. In this section, we focus on these properties and give examples of how they can be used to change meaning.

2.3.1.1 *Shape*

When observing a graphic object, shape and colour seem to have the most immediate impact. Wheeler (2009, p. 52) points out that in terms

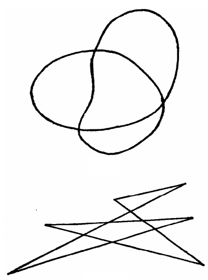


Figure 2.6: Maluma and Takete.

3 oxfordlearnersdictionaries.com/definition/english/semiotics

of sequence of visual perception and cognition, the brain firstly acknowledges shapes and only after considers colour. This is supported by perception theories (described in previous sections), which highlight the role of shape.

If we focus on shape, its visual qualities may cause us to automatically perceive an object in a certain way, even unrelated to a possible concept being represented. For example, consider the two simple shapes in Fig. 2.6. These "nonsense" shapes were firstly presented by Köhler (1929), one being described as smooth and curved, while the other as rough and jagged.

According to studies done by Lyman (1979), there are concepts consistently attributed to these seemly meaningless shapes. Some concepts are even agreed upon by 100% of participants, such as "calm" for the top shape and "angry" and "resentful" for the bottom shape. A tendency can also be observed in some abstract non-emotional concepts – e.g. 87% attribute the top shape to "eternity"; 80% attribute the bottom shape to "consciousness" (Lyman, 1979).

This attribution tendency goes one step further when considering names without any apparent meaning. According to Köhler (1929), most people attribute "maluma" to the top shape and "takete" to the bottom one, without hesitation. Ramachandran and Hubbard (2003) later studied this effect with similar shapes using the names "bouba" and "kiki". Their results show that 98% of all respondents assign the name "kiki" to the angular shape and "bouba" to the one with gentle curves. According to Ramachandran and Hubbard (2003), these name-shape attributions can be explained by our tendency to perform mappings among domains, namely between image and sound. Sharp shapes tend to be associated with "harsh-sounding" names and organic shapes with smooth ones.

Kennedy (1982) also makes reference to the change of meaning based on shape, stating that a simple change of curvature can suggest different meanings – an overbearing weight when applied to pillars or fear when applied to the legs of a human figure. Similarly, the shape of speech balloons used in comics adds meaning, often used to convey the character of sound.

2.3.1.2 *Colour*

Colour can also be used to change the meaning of a given graphic object. By simply assigning a different colour to a given element, its perceptual meaning may also change.

By analysing a simple example, it is easy to understand the importance of this aspect. Figure 2.7 shows how colour changes the perception of a banana. Consider the following three cases:

- A **green banana** is most likely unripe and not ready to be eaten;

- A yellow banana is already ripe;
- A **red banana** will probably seem strange to anyone unfamiliar with them.

Interestingly, the meaning of colour also depends on the context. Imagine a person at a traffic intersection with a traffic light (see Fig. 2.8). If they interpret the colours in a similar way to the banana example, they would probably consider the colour green as a sign for "patiently wait", which would most likely grant them a couple of honking sounds. In the same way, upon facing a red traffic light, there is often not much time to contemplate its meaning (as one might have with a red banana), because that confusion and hesitation could cause a life-threatening situation.

Figure 2.7: Bananas in different colours

This shows that choosing a colour is no ordinary task as it is directly connected to meaning. In any case, the use of colour has already been shown as a means of facilitating the interpretation – e.g. using colours that are related to what is represented improves our reading speed (Lin et al., 2013). However, the incorrect use of colour has the opposite effect. Using an unsuitable or incompatible colour may lead to interference in its interpretation, e.g. *Stroop effect* (MacLeod, 1991).

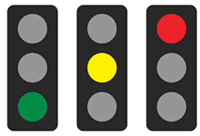


Figure 2.8: Traffic lights

2.3.1.3 *Position and Arrangement*

A third aspect that we want to mention concerns how elements are positioned. Like colour and shape, arrangement of elements is also one of the main ways of achieving specific meanings. This is observed in Fig. 2.10, in which the same elements (two hands and a card) are positioned differently, leading to three different interpretations: *give*, *receive* and *take*.

Position is often explored in visual languages as a way to represent different ideas. One example is shown in Fig. 2.9, which depicts some symbols of *Blissymbols* by Bliss (1965). By putting an arrow next to the *water* symbol a new concept is represented. The concept also varies according to the positioning of the arrow.

2.3.2 *Signs and Symbols*

One problem with the *Semiotics* definition provided at the beginning of Section 2.3 is that it mixes *sign* and *symbol*. For clarity, semiotics is normally viewed as the study of signs (Chandler, 2007) and "symbol" as one type of *sign* (as we will describe later).

A sign is considered to be the combination of a *signifier* and a *signified*. For example, when the word "cat" is read, it evokes the *mental image* of a *cat*. In this case, the word "cat" is considered the *signifier* and the mental image the *signified*. The proposal of this dyadic approach is attributed to De Saussure (2011). A different approach is proposed by

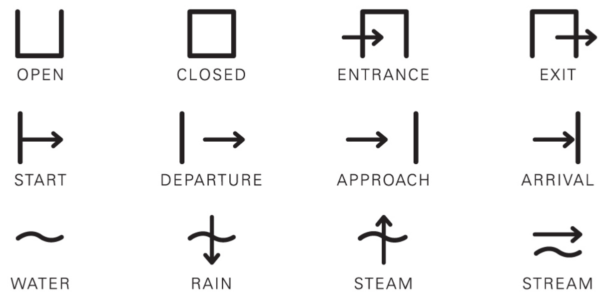


Figure 2.9: Element position in *Blissymbols* by Bliss (1965). By using the same elements in a different position, a new meaning is obtained (see symbols *start*/*departure*/*approach*/*arrival* or *water*/*rain*/*steam*/*stream*).

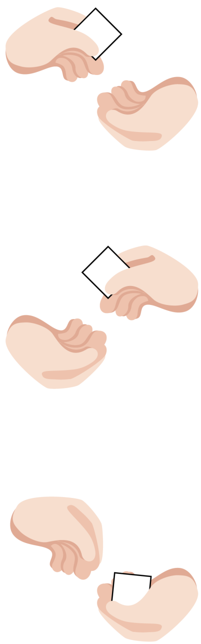


Figure 2.10: Give, Receive and Take. Adapted from Krampen, Götte, and Kneidl (2007).

Peirce (Chandler, 2007), involving three elements: *representamen* (the form taken by the sign), *interpretant* (the mental concept that is activated in the mind of the person who encounters the representamen) and *object*. In the cat example, the written word "cat" is the representamen, the mental concept of *cat* is the interpretant and a real-life cat is the object.

A simpler perspective consists in distinguishing *what is shown*, e.g. a word or drawing, from *what is meant*, a mental concept or/and a real-world object (Engelhardt, 2002).

2.3.2.1 *Types of Signs*

The work by Peirce is currently seen as one of the foundations of Semiotics. According to Peirce, there are three types of signs: *icon*, *index* and *symbol*. These can be described as follows (Chandler, 2007):

- **Icon:** a sign that bears a resemblance to the object that it represents (the *signifier* resembles the *signified*). For example, a portrait has an iconic nature.
- **Index:** a sign that connects the *signifier* to the *signified* not based on resemblance but on a relation of some sort (e.g. co-occurrence, temporal sequence, cause and effect, etc.). For example, smoke is an index of fire and a weather vane is an index of wind.
- **Symbol:** a sign that is based on an arbitrary or purely conventional association between the *signifier* and the *signified*. For example, traffic lights have a symbolic nature.

In spite of this distinction between what Peirce calls "types of signs", these are not mutually exclusive and a given sign can be seen as any combination of *icon*, *index* and *symbol* (Chandler, 2007; Engelhardt, 2002). For example, a sign that depicts a *pair of scissors* may be used as an icon when referring to scissors, as an index when referring to the act of cutting and a symbol when used to refer to the act of digitally cutting. In this sense, we are dealing with modes of relationship between *signifier* and *signified* rather than types of signs (Chandler, 2007).

A common distinction is made between *iconic*, understood as bearing resemblance to what is represented (aligning *iconic* with both literal and pictorial), and *symbolic*, as abstract and based on convention (Engelhardt, 2002). This distinction, however, is not fully accurate as *iconic* signs do not always stand for what they depict and *symbolic* ones are not always non-pictorial. For example, a pictogram of a glass (Fig. 2.11) used to represent *bar* would not be *iconic* because it does not stand for the object depicted, nor *symbolic* as it has a pictorial nature. This sort of perspective treats two dimensions – *type of correspondence* and *mode of expression* – as a single one. Nonetheless, it is important to distinguish between these two different dimensions.

2.3.2.2 *Type of Correspondence*

The way a sign is seen depends primarily on the way it is used (Chandler, 2007, p. 45). Such a view is in alignment with the distinction between two layers of meaning (Van Leeuwen, 2001): *denotation* and *connotation*. Denotation concerns what is depicted and connotation is related to the ideas and values that are expressed through what is represented (what it stands for).

Engelhardt (2002) refers to *types of correspondence* as based on the type of relationship between what is shown and what is meant. The following *types of correspondence* can be identified:

- **literal:** what is shown resembles what is meant (physical object).
- **metaphoric:** based on an analogy between what is shown and what is meant. Metaphoric correspondence may be based on structural analogy, comparable functions or shared characteristics.
- **metonymic:** based on a relationship of physical involvement between what is shown and what is meant (e.g. what is shown "is part of" or "a result of" what is meant).
- **rebus-based:** based on a similarity between the sound of the spoken word of what is shown and the sound of the spoken word of what is meant.
- **arbitrary-conventional:** what is shown stands for what is meant by convention. An example of arbitrary-conventional symbols are written words.

Engelhardt (2002) compares the *types of correspondence* with the *types of signs* proposed by Peirce (*icon*, *index* and *symbols*). Some signs identified as icons are *literal* and others *non-literal*, signs that are referred to as index (at least a subset) can be seen as metonymic, and symbols can

Figure 2.12: Encoding examples. Based on (Horton, 1994) using *OpenMoji*.

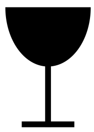


Figure 2.11: Glass

be considered as having an arbitrary-conventional type of correspondence. A more general distinction is made between literal and non-literal (*metaphoric*, *metonymic* and *arbitrary-conventional*) by Richards (1984). For a comparison between different terminology of type of correspondence, we refer to Engelhardt (2002, p. 115).

The idea to retain here is that the *type of correspondence* depends on how the sign is used. For example, an illustration of a glass (Fig. 2.11) may be *literal* if it is used to represent a cup, *metaphoric* if its intended meaning is "fragile" and *metonymic* if it is used to indicate a bar.

Moreover, the process of conveying a given meaning may be based on a sequence of references, each with its own *type of correspondence*. The pair of scissors example is exactly such a case. A computer icon that depicts a pair of scissors linked to the action of removing elements is based on a metaphoric relation between removing and the act of cutting, which in turn involves a metonymic relation between the act of cutting and an object that can be used for that purpose, i.e. scissors (Engelhardt, 2002).

In addition, different *types of correspondence* may be simultaneously used in a given graphic object. For example, a green face to denote a "nauseated person" uses a face in a *literal* way and the green colour in an *arbitrary-conventional* way (a person does not turn green when nauseated).

Focusing on the design of symbols, Horton (1994) addresses how a given meaning can be visually represented. The importance of a strong association between the symbol and the desired meaning is highlighted, which should exist not only in the mind of who designs the symbol but also in the viewer's. Horton (1994) provides a list of ways of encoding meaning in icons. We summarise the most important topics below, which are illustrated in Fig. 2.12:

- **direct representation of the subject:** there is a physical similarity between the real-world subject and the representation. This consists, for example, in the literal representation of an object (e.g. "floppy disk") or an action being performed (e.g. "exercise").

- **representation of a related or associated subject:**[4] we can represent a related or associated subject that calls the originally intended subject to mind, which is useful when the subject itself cannot be shown. Several types of association exist:
  - **logical analogies** show an analogous subject with parallels to the original one. Several types are presented: (i) *causal connection* (e.g. "maintenance" via the tool used to perform the action), (ii) *functional analogy* (e.g. "attach" via the tool used for analogous activity or "cooperate" via the body part used in activity), (iii) *results* (e.g. "center") and (iv) *structural analogy* ("digital link").
  - **metaphors and figures of image** are used to express something that cannot be literally represented. Several types are presented, among which: (i) *synecdoche* (part for the whole, e.g. "restaurant"), (ii) *litotes* (representation by negation, e.g. "peace"), (iii) *hyperbole* (exaggeration, e.g. "fatal error"), (iv) *euphemism* (substitution for a more acceptable subject, e.g. "injury").
  - **conventions** consist in using a symbol that is already considered conventional (e.g. "radioactive hazard").

This list of ways to encode meaning is aligned with the *types of correspondence* identified by Engelhardt (2002).

2.3.2.3 *Mode of Expression*

Graphic objects can be classified in terms of *mode of expression*. Although there are differences in terminology among authors,[5] we consider two different *modes of expression*:

- **pictorial:** depicts a physical object or scene, and is positioned on a rendering spectrum from realistic to schematic;

- **non-pictorial:** does not depict a physical object. Examples are an abstract shape, a word or a number.

Despite the need for distinguishing *mode of expression* from *type of correspondence*, there is an obvious connection between the two: a literal type of correspondence is based on a pictorial mode of expression. As such, as soon as a graphic object is interpreted as involving literal correspondence it should be considered as pictorial. Interestingly, a given object may have different classifications depending on what it is used

4 Horton (1994) uses the term "object", which in our opinion is too limited.
5 We refer the reader to Engelhardt (2002, p. 122) for a detailed comparison between different terminologies.

Figure 2.13: Spectrum of Abstraction. Adapted from McCloud (1993).

for. For example, a circle can be used in a non-pictorial way – e.g. used in a Venn diagram – or in a pictorial way – e.g. a circle to represent *moon*. Moreover, a pictorial use can also be interpreted as non-literal – e.g. a circle meant as a moon to represent *night*.

A note should also be made regarding the distinction between *pictogram* and *ideogram*, which depends on the nature of the graphic object depicted. A *pictogram* (also referred to as *pictograph*) is a graphic object that has a pictorial resemblance to a physical object, for example the glass pictogram (Fig. 2.11). On the other hand, an *ideogram* (also referred to as *ideograph*) is a graphic element that cannot be fully related to a resemblance to a physical object. For example, the "Do not drink" symbol shown in Fig. 2.14, which combines the glass pictogram with a prohibition symbol.

Figure 2.14: "Do not drink" symbol

Regarding pictorial objects, their rendering may vary in terms of degree of abstraction (Fig. 2.13), from very *realistic* to very *schematic*. This variation occurs regardless of the intended type of correspondence. In Fig. 2.13 two versions of a face are shown, from a more realistic to a more schematic. An interesting aspect is that the simpler and more schematic it is, the more people it can represent.

This interplay along the spectrum of abstraction is also related to perception and convention. Consider, for example, the representation of a heart with an arrow, which is conventionally associated with love (Cupid's arrow hitting the heart). A usual representation for it is the illustration in the middle of Fig. 2.15, which uses the symbol (ideogram) that is anatomically inaccurate but conventionally associated with heart. A more realistic version of it (top on Fig. 2.15), using a literal representation of a heart, would likely be considered strange and unusual. On the other hand, using a more abstract version (bottom on Fig. 2.15) would difficult its interpretation.

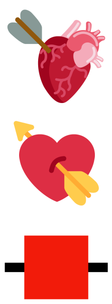


Figure 2.15: Heart with arrow.

An abstraction process can be seen as focusing on specific details rather than eliminating details – i.e. "stripping down an image to its essential 'meaning"' (McCloud, 1993). This is related to the reduction to essential parts as a way to represent concepts, mentioned in Section 2.2.2. The process of reduction through abstraction has been explored multiple times in the History of Art. Consider, for example, Pablo Picasso's *The Bull*, a set of eleven lithographs produced in 1945 based on a sequential abstraction (Lavin, 1993). Another interesting example can be seen in the illustrated books by the Swiss illustrator

Warja Lavater, which re-tell classic fairy tales through a symbolic use of shape and colour (e.g. Lavater, 1965).

2.3.3 *Interpretation*

One last aspect that we would like to mention concerns interpretation. When producing a visual representation, its correct interpretation is not guaranteed. A classic example is given in the book *The Little Prince* by Saint-Exupéry (1943): the main character presents a drawing to viewers, who interpret it as a hat, failing to perceive its intended meaning – a boa constrictor digesting an elephant. The matter of interpretation is especially relevant in the case of representations with abstract expression and a non-literal type of correspondence. An example can be observed in Fig. 2.16, which represents a giraffe passing behind a window. However, the interpretation may be hard if one is not aware of this intended meaning.

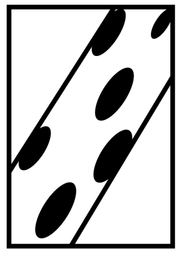


Figure 2.16: Abstract shapes or something more? Adapted: (Arnheim, 1974a).

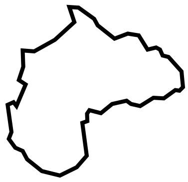


Figure 2.17: Rotated Africa. Adapted: (Zimbardo and Gerrig, 2002)

Even though a given visual representation is designed with a given meaning as a goal, its interpretation always depends on prior knowledge – as stated by Arnheim (1974a, p. 48), what a person sees is the outcome of what they have seen in the past. As such, our perception of the word relies on our past experiences and on the knowledge derived from them. Meaning can then be seen as a social construct, which is neither fixed nor absolute but rather context- and culture-dependent (Engelhardt, 2002). For example, the shape presented in Fig. 2.17 depicts Africa but it may be difficult to immediately perceive it as such, due to the conventionalised (unrotated) version. Another example is the dustbin as a symbol for deletion (Ware, 2012). If one is not familiar with dustbins and their purpose, they do not possess the necessary knowledge to make this metaphoric interpretation. Moreover, for the same concept, different cultures may have different representations for the same concept, e.g. the "Red Crescent" is used in Muslim countries instead of the "Red cross" (Hassan, 2015). These aspects can be summarised as an interplay between three main factors, represented in the following equation (Horton, 1994):[6]

$$image + context + viewer = meaning. \tag{2.1}$$

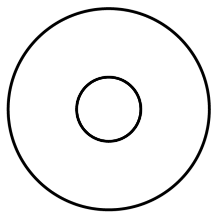


Figure 2.18: Mexican with sombrero viewed from above.

Lastly, one graphic object may have several possible interpretations. This aspect was already mentioned when we described the different *types of correspondence* (e.g. *literal* vs *metaphoric*), which depends on how the object is used. However, even just considering a literal interpretation, an object may be considered to represent different things. One example can be observed in Fig. 2.18, which can be interpreted as a "Mexican wearing a large sombrero viewed from above", a "millstone" and even a "doughnut" (Arnheim, 1974a).

6 Horton (1994) uses "icon" but we generalise to "image".

2.4 VISUAL GRAMMAR

In the previous sections, we have described the set of *visual properties* that are available for producing visual artefacts, how these can be used to convey meaning and ultimately allow one to represent concepts. In this section, we focus on how elements are put together to construct more complex visual objects. Engelhardt (2002) studies this topic and summarises the different aspects that it involves:

> *Part of what a graphic representation means depends upon the graphic objects that it contains, and part depends on the graphic relations that those graphic objects are involved in.*
> (Engelhardt, 2002, p. 12)

This quote is inspired by a language-related one, originally stated by Minsky (1985) in reference to syntax in language. With this appropriation, Engelhardt (2002) makes a parallelism between written/spoken language and the visual domain, in an attempt to establish a foundation for what is referred to as *Graphic Syntax*. As in spoken and written language, syntax is the part of grammar that concerns order and structure. Horton (1994) defines grammar as the set of rules for combining symbols, whether the symbols are words or pictures. Several authors have proposed the notion of "Visual Grammar" (e.g. Leborg, 2006), while others have criticised such idea (e.g. Forceville, 1999). In any case, it is unlikely that a set of universal visual grammar rules exist, to which all visual systems abide. Despite this, the design of certain types of images seems to follow structuring guidelines, e.g. *data visualisations* (Engelhardt, 2002), *comics* (Cohn, 2013; Walker, 2000) and *icons* (Horton, 1994).

Coming from the context of data visualisation, Engelhardt (2002) proposes a framework for the analysis and decomposition of the visual language of graphic representations. Their approach is based on the understanding of graphic representations as graphic objects, which have a composition that involves the use of visual attributes (e.g. *colour*) and relations among elements and space (e.g. *labelling*, *superimposition*, etc.), and the existence of different syntactic roles. We draw from the work by Engelhardt (2002) to identify aspects to consider when following a syntax-based approach for the production visual representation of concepts.

2.4.1 *Visual Structure*

As Engelhardt (2002) frames it, part of the meaning of a graphic representation depends on the graphic objects that it contains, and part depends on how those graphic objects are arranged and on the relations that the objects are involved in. This topic was briefly mentioned in Section 2.3.1, by providing examples of how arrangement affects meaning.

The graphic object is seen as a carrier of *visual attributes*, such as *size*, *shape* and *colour*. Despite this, a graphic object is often equated with its shape, which is regarded as the carrier of other visual attributes (e.g. the large red square). Moreover, Engelhardt (2002) describes the analysis of a graphic representation as a recursive process, due to the fact that a given graphic object can be a nesting of simpler graphic objects. The objects at the lowest level of decomposition are referred to as elementary graphic objects. In this sense, a graphic object can either be *elementary* or *composite*. A *composite graphic object* is composed of:

- a graphic space (that it occupies);
- a set of graphic objects contained within the graphic space;
- a set of graphic relations in which the graphic objects are involved.

Even though this approach was proposed with the analysis and decomposition of visual representations as a goal, we argue that it can be used for the construction of new visual representations. In the following sections, we will address two other aspects mentioned contemplated in the framework proposed by Engelhardt (2002): *relations* and *syntactic roles*.

2.4.2 *Relations*

Engelhardt (2002) identifies two *types of relations*: *spatial* and *attribute-based*. In addition to this distinction, relations can also be categorised based on the elements involved: *object-to-space* and *object-to-object*.

An example of an attributed-based object-to-object relation is "[x] has the same colour as [y]". This sort of relation is key for the creation of patterns and establishing a perception of similarity or distinction, which are key in data visualisation (Bertin, 2011).

Regarding object-to-space relations, they are one of the two ways that spatial structure can be created. They concern the positioning of objects in the metric space. This type of relation is central in the context of data visualisation (e.g. for the production of diagrams, networks and maps) but less important for purpose of this thesis (symbols).

The other way of creating spatial structure is by resorting to spatial object-to-object relations, which can be seen as types of object-to-object "anchoring". Engelhardt (2002) proposes a total of six types of spatial object-to-object relations (see Fig. 2.19):

- *spatial clustering*: arrangement of a set of graphic objects into two or more groups via within-group proximity;
- *lineup*: arrangement of graphic objects in a line;
- *linking by a connector*: connection between graphic objects using a connector (e.g. an arrow);

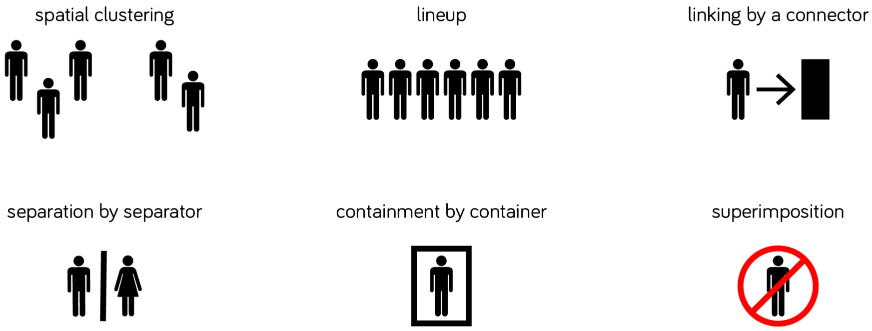


Figure 2.19: Spatial Object-to-object by Engelhardt (2002)

- *separation by a separator*: separation of graphic objects through the use of a separator (e.g. a line);
- *containment by a container*: one graphic object contains others by visually surrounding them;
- *superimposition*: a foreground object is perceived as being in front of the background object, occluding parts of it.

Some of these relations are in line with *Gestalt* principles – e.g. *spatial clustering* is based on *proximity*. However, we argue that these six types are not on the same level of concreteness, some being more specific than others. For example, *spatial clustering* only makes reference to the existence of groups build via *proximity*, whereas *containment* is more specific as it depends on the existence of a *container element* and at least one *contained element*.

Leborg (2006) presents what is referred to as *relations*. Some examples are *attraction* and *parallel* (the full list is shown in Fig. 2.20). Similarly to what was said of the relations proposed by Engelhardt (2002), these can also not be said to be fully equivalent on concreteness. For example, *direction* only hints at the existence of a sense of direction, defined as "a structure can actively define a direction" (Engelhardt, 2002, p. 62). In contrast, *tangent* is clearly defined as two objects located next to each other and sharing a common point (Engelhardt, 2002, p. 76).

In general, we consider that these two lists of relations – spatial object-to-object by Engelhardt (2002) and the ones by Leborg (2006) – are excessively high level and, consequently, difficult to use in the production of graphic representations.

On the other hand, other perspectives are based on *low-level relations*. Ungerer and Schmid (2006) make reference to *locative relations*, such as *over*, *under*, *up*, *down*, *in* and *out*. These are characterised as *orientational image schemas*, which are derived from our bodily interaction with the world. As Ungerer and Schmid (2006) points out, these are not understood as abstract principles but as mental pictures, thus being easy to use in pictorial representation. Engelhardt (2002) also makes a reference to these *image schemas* when comparing their terminology with

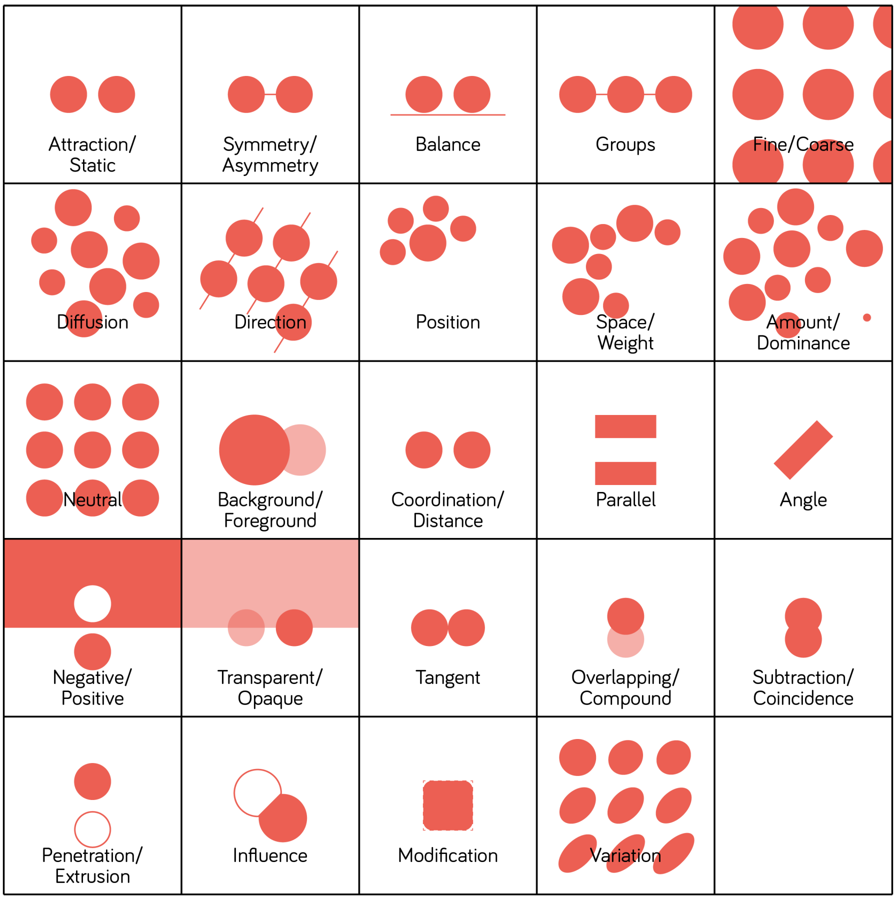


Figure 2.20: Relations by Leborg (2006) (adapted).

the one by Lakoff (1990), who addresses the notion of *kinesthetic image schemas*, e.g. UP-DOWN and CONTAINER.[7]

In fact, insight can be drawn from research focused on the learning of spatial prepositions. Regier (1996) uses connectionist models to explore the learnability of spatial prepositions from several languages, using positive and negative schematic examples. In English, they use the set *above*, *below*, *left*, *right*, *around*, *in*, *on*, *out of*, *through* and *over*. Similarly, Wang et al. (2008) focus on how humans describe relative positions of objects and identify the three most used types of relations: *directional* (*right*, *left*, *above* and *below*), *topological* (e.g. *overlap*, *separate*, *touch*, *in*, *out*, etc.) and of *distance* (e.g. *near*, *far*, etc.).

These types of relations have been used in the implementation of computational systems for the description of objects in images (e.g. Falomir et al., 2012; Roy, 2002). As such, they seem appropriate as object-to-object spatial relations between graphic elements for the purpose of visual representation.

In Section 6.2.1.2, we present a list of visual relations that we used in our experiments with the *Blender*.

7 For a full terminology comparison, we refer the reader to Engelhardt (2002, p. 52).

2.4.3 *Combination and Syntactic Roles*

Going back to the set of *spatial object-to-object relations* proposed by Engelhardt (2002), there are two relations of special interest for the purpose of this thesis: *containment*, i.e. a graphic object contains other graphic objects by visually surrounding them, and *superimposition*, i.e. a foreground object is perceived as being "in front of" a background object, occluding part of it (see Fig. 2.19).

These two types are central in the notion of what Engelhardt (2002) refers to as *composite symbol*. A *composite symbol* is defined as a special case of a composite graphic object, composed of several graphic objects that are arranged in a conventionally fixed arrangement, usually involving *containment* or *superimposition* (Engelhardt, 2002).

Figure 2.21: Addition (family), Antithesis (contrast), Overlap (tools) and Specification (music folder). Based on (Horton, 1994) using *OpenMoji*.

Horton (1994) addresses the topic of symbol combination and states that it should be done according to a set of rules, which are referred to as "grammar". In their perspective symbol combination can be used to represent more complex concepts by combining elementary symbols. They identify four basic ways of combination, which are the following:

- Addition: adding the independent meanings of separate objects, representing a concept that is the sum of its parts (e.g. family in Fig. 2.21).
- Antithesis: combining two contrasting elements in order to suggest a richer idea than what is conveyed by either alone (e.g. contrast in Fig. 2.21).
- Overlap: combining elements to represent an abstract or general concept that is the only common aspect among the elements (e.g. tools in Fig. 2.21). This can be seen as related to what Reale et al. (2021) refers to as "hyperonymic combinations".
- Specification: combining elements that restrict each others meaning (e.g. music folder in Fig. 2.21).

In addition to addressing ways of combining elements, Horton (1994) mentions that a graphic grammar should specify, among other things, which elements are required and which are optional. This is particularly important in the case of *specification*. Horton (1994, p. 184) highlights the importance of picking a *border*, as it affects the way the symbol is perceived through meaning clarifying. With *border* Horton (1994) refers to not only a square-shaped element that encloses a given symbol but also to the outline of an object. This way, the image within the border indicates a specific instance of the concept represented by the border element (see Fig. 2.23). This notion of *border* is similar to *containment* by Engelhardt (2002).

Elements involved in *containment* and *superimposition* usually take certain syntactic roles, which are based on transformation or further

Figure 2.23: Phone book. Based on (Horton, 1994, p. 186) using *OpenMoji*.

specification of the meaning. Engelhardt (2002) identifies several other *syntactic roles* (e.g. *node*, *label*, *separator*, *connector*, etc.) but we will focus our attention on the ones most relevant to this thesis.

In the case of *containment*, there are two different roles: a *container element* and a *content element* (contained). The meaning of a container-based symbol is the outcome of the combination of the elements. An example can be observed in Fig. 2.22, which shows the pictogram for "bike", the symbol for "road closed" and a symbol that is the combination of the two through containment, normally assigned with the meaning "no cycling".

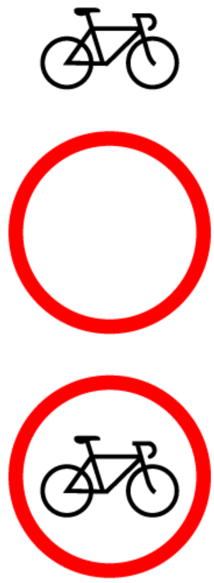


Figure 2.22: Bike, Road Closed and No cycling.

Engelhardt (2002) also mentions the role of *modifier*, in which an element is superimposed on another, changing its meaning (for an example, see *superimposition* in Fig. 2.19). Similarly, comics are considered to have a language system, in which certain elements have specific roles and can be used to modify or specify meaning (Cohn, 2013; McCloud, 1993; Walker, 2000). For example, elements, such as *droplets*, *spikes*, or *spirals*, are shown to signal an emotional state when positioned around the head of a character (Ojha, Forceville, and Indurkhya, 2021).

A correspondence can be made with linguistic analysis by resorting to the notion of *morphemes*. Morphemes are considered as the smallest meaningful components of speech, for example, the word "sleepwalk" has two morphemes "sleep" and "walk". Engelhardt (2002, p. 24) refers that the distinction between *free morphemes* (occur by themselves) and *bound morphemes* (occur attached to other morphemes) can also be made in graphic objects. In this sense, *content objects* (e.g. a cigarette drawing) could be seen as *free morphemes*, *modifiers* (e.g. a red cross to imply prohibition) as *bound morphemes*, and some *containers* as *free* (folder) while others as *bound* (blue circular in traffic sign).

Figure 2.24: Indexical line (top), upfix (middle)and eye-umlaut (bottom). Adapted from Cohn (2013).

Cohn (2013) addresses this topic and refers that visual languages use the same strategies of morphological combination as verbal languages (e.g. *prefix* as morpheme placed in front, *suffix* as placed at the end, etc.). Cohn (2013) provides a list of strategies used in comics, for example: *affixation* by using "carriers" (e.g. the speech balloons), which are related to the notion of *container*), *indexical lines* (e.g. lines to indicate motion, see Fig. 2.24) and *upfixes* (e.g. elements placed above a head, see Fig. 2.24); and *suppletion* (one morpheme fully replaces another) by using eye-umlaut (replacing the eyes with elements, see Fig. 2.24). Recently, Cohn and Foulsham (2022) analyse conventional and unconventional face-upfixes in comics and emoji.

本 木 林 森

Figure 2.25: Chinese characters for *root*, *tree*, *woods* and *forest* (left to right). *Root* can be obtained by adding a line to the *tree* character; *woods* character can be obtained by combining two *tree* characters; *forest* can be obtained by combining three *tree* characters (Yan et al., 2013).

Overall, these syntactic roles are considered to be crucial for the production of *composite symbols* (Engelhardt, 2002). Moreover, it is often the case that several composite symbols share a given visual vocabulary and a compositional grammar, being considered as belonging to a "family". Returning to the example shown in Fig. 2.22, the symbols are part of the "traffic sign grammar", in which it is often the case that a symbol uses a container object related to a specific meaning (e.g. permission or prohibition) and a content object (e.g. car). In the next section, we will give some examples of visual language systems.

2.4.4 *Visual Language Systems*

The aspects described in the previous sections are the foundation for the development of *visual language systems*. As stated by Reale et al. (2021), writing evolved from expression of the meanings to the expression of the sounds. The earliest writing systems were composed of *logograms*, which represented concepts. Later, there was an evolution towards *phonographic systems*, in which characters were used to represent sounds instead of concepts. Nonetheless, there are still writing systems that employ both *phonographic* and *non-phonographic* strategies (e.g. Chinese).

Chinese writing is particularly interesting for its multiple categories of characters (Hew et al., 2012; Qiu, 2000). For example, some characters have a *pictographic* nature, for example the character used to represent "tree" (see Fig. 2.25). These characters belong to the formation category known as *Pictogram Chinese Characters* and are the result of an evolution process from more *pictorial* to more *abstract*. Other characters are considered *ideograms*, for example the character for "root" is produced by adding a line to the bottom of "tree". Some characters are categorised as *Ideogrammic Compounds Chinese Characters* and can be decomposed into others, revealing that the concepts behind them are semantically related, for example "woods" is composed of two trees (see Fig. 2.25). Despite this, most Chinese characters are actually *phono-semantic*, combining phonetic with semantic.

More recently, several *non-phonetic* languages have been developed, in which signs express meaning instead of sounds – these are called *semasiographies*. The development of these languages was motivated by multiple reasons, among which the dream of a universal language (Horton, 1994). Reale et al. (2021) identify several non-mutually exclu-

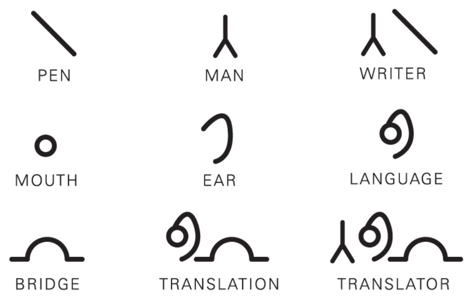


Figure 2.26: *Blissymbols* by Charles Bliss

sive categories into which existing semasiographies can be grouped. In the following paragraphs, we present the most relevant categories by giving examples of existing semasiographies.

One of the best known visual language systems was proposed by Bliss (1965), who developed a communication system composed of several hundreds of ideographs that can be combined to make new ones – *Blissymbols*. Several things can be observed by looking at *Blissymbols*: such as a variation in terms of abstraction degree (there are both *pictorial* and *abstract* symbols); by combining symbols, new meanings are obtained (examples in Figure 2.26: *pen* + *man* = *writer*, *mouth* + *ear* = *language*); by using the same symbols in a different position, a new meaning is obtained (see symbols *water*/*rain*/*steam*/*stream* in Fig. 2.9). *Blissymbols* have been used as a method of Augmentative and Alternative Communication (AAC) since the 1960s (Lin and Biggs, 2006).

The first category of semasiographies is precisely *iconic languages used for AAC*, which are used to facilitate communication for people with disabilities. Another example is Picture Communication Symbols (PCS), which has as main goal to provide symbol-based products, training and services for individuals with special needs. When compared to other pictogram-based systems, its symbols are more colourful and representational (giving preference to human expressions over abstract graphics, e.g. a smiley face instead of a heart to represent "like") (Lin and Biggs, 2006).

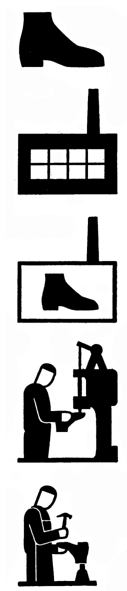


Figure 2.27: ISOTYPE examples. Adapted from Neurath (1936).

The second category by Reale et al. (2021) that we want to mention is *Art and Graphic Projects*. One example of this category is ISOTYPE (International System of TYpographic Picture Education), which was developed by Otto Neurath and Gerd Arntz with the purpose of making the developments in society clearer by using pictorial statistics (Jansen, 2009; Neurath, 1936). In ISOTYPE the first step was to develop symbols that would be easy to understand and to remember (e.g. pictogram of "shoe" and "works"/"factory" in Fig. 2.27). Then, these could be combined to represent other concepts (e.g. "shoe factory", "shoes made by machine" and "shoes made by hand" in Fig. 2.27).

Another example from the category of *Art and Graphic Projects* is *Elephant's Memory*, which was created by Timothee Ingen-Housz with the

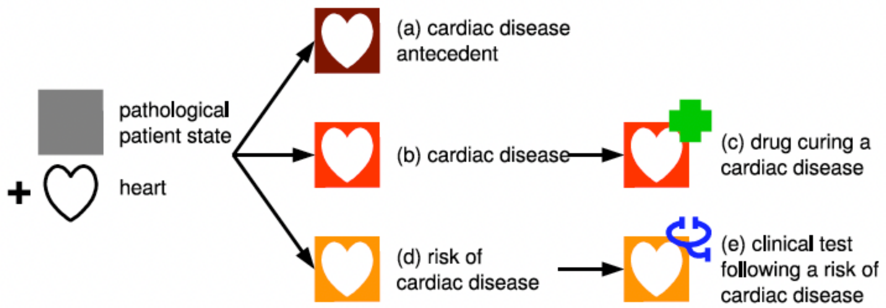


Figure 2.28: VCM examples. Source: Lamy et al. (2008).

goal of fostering new ways of envisioning communication (Lin and Biggs, 2006). One interesting aspect is that it is a non-linear language, enabling messages to be written and read from different starting points (Reale et al., 2021).

The third category that we want to mention is *Sectorial Visual Languages*, which is described as being conceived for specific purposes (understood as specific domains). One example is VCM (*Visualisation des Connaissances Médicales*; *Visualisation of Medical Knowledge*), which is an iconic language designed by Lamy et al. (2008) with the goal of facilitating the use of recommendations provided in drug monographs. The language is based on combinatorial graphical grammar and a small number of graphical primitives (see Fig. 2.28). It can be used to represent diseases, physiological states, life habits and other aspects described in drug monographs.

Finally, Reale et al. (2021) mention a category related to *Emoji*. Two examples are: the *Sitelen Emoji*, which is a project released in 2019 that assigns emoji to each word of the constructed language *Toki Pona*;[8] and *Emojese*,[9] which is a project released in 2021 that explores the construction of a written language using emoji.

Other categories have been proposed but we refer the reader to Reale et al. (2021) for a more detailed description. Reale et al. (2021) compare existing semasiographies in terms of four characteristics: *compositionality* (elements can be combined to produce a representation whose meaning is the semantic composition of the original meanings), *iconicity* (relation between the visual signifier and the signified), *universality* (not bound to a specific area) and *language independence* (not related to a specific language). In the comparison, an additional semasiography is also included: IKON. Reale et al. (2021) present IKON, which was created with the purpose of satisfying the four characteristics identified, thus improving the existing offer for visual communication.

Partially aligned with what was described in the previous sections, Reale et al. (2021) identify the most common ways of building non-phonographic characters: *pictograms* (represent what they mean), *pic-*

8 sites.google.com/view/sitelenemoji/home
9 emojese.org/

tographic scenes (combination of elements to describe a different concept), *metaphorical/symbolic signs* (using elements in a combination as metaphoric reference), *abstract signs* (concepts represented using abstract elements) and *hyperonymic combinations* (pictographic representations are combined to represent a common category).

Overall, from the analysis to the existing visual languages, one can easily observe the importance of *compositionality* (e.g. new meanings result from the combination of individual elements in *Blissymbols*) and *syntactic roles* (e.g. containers in ISOTYPE and modifiers in VCM) for the production of visual representations of concepts.

2.5 SUMMARY

In this chapter, we started by identifying *perceptual features* that can be used in the visual representation of concepts. Then we focused on concepts and described the aspects that should be considered when producing visual representations for a given concept. Afterwards, we shifted our attention to how visual features and meaning can be linked through *Semiotics*. The last part of the chapter describes topics related to visual grammar, such as the *structure* of graphic representations, *relations* among elements and different *syntactic roles*. In summary, we identify the following important aspects:

- the identification of *prototypical features*;
- considering *affordances* in representation;
- *conceptual extension* and using properties from related concepts;
- the use of visual properties to confer meaning, through *semiotics*;
- the combination of elements to achieve different meanings and the use of *syntactic roles*.

3 COMPUTATIONAL APPROACHES TO VISUAL REPRESENTATION OF CONCEPTS

In this chapter, we provide an overview of the state of the art of computational approaches to the visual representation of concepts. First, we briefly describe three research areas: *Computational Design*, *Computational Creativity* and *Computational Co-creativity*. Then, we focus on concrete computational approaches.

3.1 CORE TOPICS

This Section introduces concepts that we consider crucial to frame the work described in this thesis. We focus on three topics: *Computational Design*, *Computational Creativity* and *Computational Co-creativity*. Even though they partially overlap, it is important to distinguish them.

3.1.1 *Computational Design*

Computational Design can be directly interpreted as the process of designing using computational means. Although this interpretation is not wrong, it is too broad and encompasses many distinct subjects. First, by using the aforementioned interpretation, one has to consider the act of designing using software like *Adobe Illustrator*. This interpretation is made, for example, by Martins et al. (2019b), who refer to proprietary design software as "conventional computational design tools". Despite this, considering the use of such software as Computational Design is, in our perspective, a lesser use of the term, in the sense that, in most cases, the computer is not more responsible for the design than a pencil would be for a drawing.

A similar view is expressed by Worrall (2020) in the context of audio design and production, who highlights the distinction between using software (e.g. *Audacity*) for audio editing, mixing, etc., and taking advantage of computational power to employ algorithms in the design process. Examples of the latter are the use of real-time responsiveness to user input or environment (Richardson, 2017), the introduction of analogy-based procedures (Gero, Grace, and Saunders, 2008; Goel, 2019) or case-based search strategies (Ayzenshtadt et al., 2017), and the implementation of methods for personalised curiosity modelling (Grace et al., 2017). As such, Worrall (2020) uses the term "Computerised Design"[1] to refer to the former, which encompasses Computer-

1 An important note should be made on the term "Computerised", which is used differently by Lelis (2021) to characterise smart brands that make use of real-time data

aided Design (CAD) tools, leaving "Computational Design" to the latter, which may address topics such as *Artificial Intelligence in Computer-Aided Design*, *Knowledge-Supported Design* or *Case-Based Design*.

Similarly, Reas and McWilliams (2010, p. 25) identify two categories of using software in arts: *production*, in which a computer is used to produce a preconceived form; and *conception*, in which the computer participates in the development of the form. These two categories are aligned with the two terms by Worrall (2020). Reas and McWilliams (2010, p. 25) go on to highlight that proprietary software products are somehow limited to the production of specific types of forms. As such, if one wants to go beyond these limitations, they either need to customise existing applications through programming or produce their own software. One example of the former case is the work by Ferreira (2019), who uses scripting in *InDesign*[2] to generate different layout possibilities for the same input content.

Another perspective is the one by Yu, Gu, and Ostwald (2021, p. 2), who state that computational design can both be defined as a model of the design process – conceptualised in computational terms, considering it as an iterative operation with *inputs*, *rules* and *outputs* – and a consideration of the tools used to support it. For the context of this thesis, the focus is on the latter. Yu, Gu, and Ostwald (2021, p. 2) make a note of the fact that the term "computational" refers to the use of logic or algorithmic systems, without the actual need to involve a computer. Nonetheless, Yu, Gu, and Ostwald (2021, p. 11) view computational design as including both CAD tools and the use of programmatic and generative approaches to automating design, thus going back to the initial interpretation.

3.1.1.1 *Generative, Parametric and Algorithmic Design*

A couple of terms are often used interchangeably in regards to computational approaches to design automation, such as "generative design", "parametric design" or "algorithmic design" (Agkathidis, 2015, p. 8). Despite being similar, they do not mean exactly the same. *Parametric design* is considered by Yu, Gu, and Ostwald (2021, p. 21) as a specific type of *generative design*. It can be seen as a recipe-like approach, in which values (*parameters*) are given as *input* and have an effect on the *output* of the design process (Reas and McWilliams, 2010). *Generative design* is seen by Yu, Gu, and Ostwald (2021, p. 21) as a broad category that covers computational methods or algorithmic approaches that are used to automate the process of producing design solutions.

and programming to implement transformations to their visual attributes or elements (colour, shape, etc.).

2 Despite being proprietary products, *Adobe's Illustrator*, *InDesign* and *Photoshop* have been complemented with the possibility of using scripting, which enables users to extend the tools through code writing (Reas and McWilliams, 2010, p. 31).

Algorithmic design refers to design approaches that involve the use of algorithms, which overlaps with the other two terms.

As Yu, Gu, and Ostwald (2021) mention, both *parametric* and *generative design* involve the task of designing with algorithms. A note should again be made on the fact that not all generative approaches use a computer. Agkathidis (2015) mentions this by stating that several generative techniques existed long before the digital revolution. Martins (2021, p. 11) supports this view and summarises the goal of *generative design* as the creation of an algorithm to produce multiple designs for a particular concept, which can be executed by either people or machines.

To highlight the potential of *generative design*, Martins et al. (2019b) state that "conventional" computational design tools are insufficient when it comes to exploration in the initial stages of the design process, limiting and biasing the designers – in line with the aforementioned views by Reas and McWilliams (2010). Martins et al. (2019b) argue that the designer should exploit the tools at hand by modifying them or inventing new ones, thus creating their own tools. The generative approach is seen as a solution, making the designer move away from making decisions about a single solution to designing the process that produces an array of possible designs (Reas and McWilliams, 2010, p. 93). These approaches have made it possible for a designer to change its method from producing and analysing a few design solutions to being able to quickly generate multiple design options and use computational methods to select the best ones (Yu, Gu, and Ostwald, 2021, p. 4). Gross et al. (2018, p. 3) frame it as "the designer shifts from being a performer of tasks to being a conductor, effectively orchestrating the decision-making process of the computer".

Examples of generative design can be seen in the context of *Dynamic Visual Identities* (Guida, 2014), in which generative approaches are used to make identities react to external data through variation mechanisms (Martins et al., 2019c) – e.g. the Nordkyn visual identity was designed to change according to meteorological data. Other approaches involve the use of generative grammars (Stiny and Gips, 1971), which employ combination rules to generate objects, e.g. houses in the style of the architect Siza Vieira (Duarte, 2005).

Yu, Gu, and Ostwald (2021, p. 26) categorise generative systems into four non-exclusive categories: (i) *generative grammars*, which use transformational rules that can be applied recursively to develop an object or shape; (ii) *evolutionary systems*, in which a recursive process of design reproduction takes place, considering a solution as an "organism" that can be used to produce offspring; (iii) *self-organising emergent systems*, in which the outcomes emerge out of self-organised components; and (iv) *associative generation*, which is based on the establishment of relationships between different components, thus making it so that any change made to their properties will lead to the generation of new instances (e.g. *parametric design*).

For this thesis, the most relevant category concerns *evolutionary systems*. As such, we will address it in more detail in the next section.

3.1.1.2 *Evolutionary Design*

One of the challenges in *generative design* concerns the exploration of the design space to find interesting solutions. Due to the nature of the generative approaches, the search space is often too big for the designer to explore on their own. A common strategy involves the use of Evolutionary Computation (EC), which is a subfield of Artificial Intelligence (AI) in which search algorithms are inspired in biological evolution, such as the Theory of Natural Selection by Darwin (1859). The use of EC in Computational Design is often referred to as *Evolutionary Design* (Bentley and Corne, 2002). An early example are the evolution of artificial creature-like forms called biomorphs by Dawkins et al. (1986).

Briefly described, a common EC strategy (known as *Genetic Algorithms*) involves the evolution of individuals (candidate solutions) in a population (Eiben and Smith, 2015). First, an initial population of individuals is generated. Then, solutions are evolved by evaluating individuals, according to a given fitness function that assesses the quality of the solution. Afterwards, the best solutions are selected and used to generate offspring, producing the next generation of individuals.

The fitness assignment depends on the problem and different approaches can be used, either based on automatic fitness assignment or through user interaction. This latter approach is commonly known as Interactive Evolutionary Computation (IEC) (Takagi, 1998) and consists in allowing users to select individuals in an interactive manner.

In this thesis, we focus on computational design as the act of using programming to implement computational systems that can be used to design. In the developed projects, we explored EC strategies, in particular resorting to interactive evaluation, i.e. IEC. One important aspect is mentioned by Martins (2021, p. 12), who points out that the randomness implemented in generative design systems provides unpredictability to the design process and to its outputs, which can promote the creativity of the designer. On this subject, a question can be raised related to creativity: is creativity something reserved to the human author, eventually fostered by a system, or can a computational system also be described as creative? To address this question, we introduce the topic of *Computational Creativity*.

3.1.2 *Computational Creativity*

Computational Creativity (CC) can be defined as a subfield of AI, focused on the study of computers as autonomous creators and co-creators (Veale and Cardoso, 2019, p. 2). Research conducted in the field of CC addresses problems from many different domains (Loughran and O'Neill, 2017), among which games (e.g. Guckelsberger, 2020), museology (e.g. Long, 2021), musical composition (e.g. Bodily, 2018), poetry (e.g. Lamb, 2018) and storytelling (e.g. Wicke, 2021).

In the previous section on *Computational Design*, we addressed the topic of generative design, in which the focus is on the production of multiple outputs for a given design problem. When entering the field of CC, one key aspect that is often given attention to is the distinction between mere generation, in which *quantity* and *novelty* do not imply *value*, and what is said to be true creativity (Ventura, 2016).

One of the central questions in CC concerns exactly its definition ("what is Computational Creativity?"), which is intrinsically related to the definition of creativity itself. Different views exist on the topic but the work by Boden (1998, 2004) can be considered as one of the foundations for the work developed in CC. First, Boden (2004) makes the distinction between two types of creativity: "psychological" or "personal" *p-creativity*, in which an individual comes up with something, e.g. a product, that is novel for themselves, and "historical" *h-creativity*, in which the same developed product has never been created before by anyone, thus considered historically novel. Second, Boden (1998) proposes a framework for the study of creativity in AI, describing the three forms of creativity used in the creative process: *combinational*, based on the novel combination of existing ideas; *exploratory*, based on the generation of novel ideas through exploration of structural conceptual spaces; and *transformational*, based on the transformation of a given dimension of the conceptual space, allowing new structures to arise.

A popular definition of CC is proposed by Colton and Wiggins (2012). Colton and Wiggins (2012) considers computational creativity systems as those that exhibit behaviours that unbiased observers would consider creative. In such formulation, they provide a working characterisation of CC without the need of defining creativity. However, their formulation makes the assessment of computational creativity dependent on an observer, who only needs to perceive the system as creative to make it so (Jordanous, 2012).

Overall, the notion of a creative process is linked to the production of *novelty* and *value* (Boden, 2004; Wiggins, 2006). Aligned with this, Ventura (2016) describes what is referred to as a spectrum of candidate computationally creative processes, based on the presence (or lack thereof) of three characteristics: *novelty* (quality of the product of being new), *value* (the worth or usefulness of the product) and *in-*

tentionality (the system's capability of having a goal). These characteristics are proposed as indicators of creative behaviour.

Pérez (2018) characterises the work developed in the field of CC as a continuum, with one pole being related to an engineering-mathematical perspective and the other to a cognitive-social perspective, in which creative agents can be positioned according to their goals as systems. This view is in line with the goals of CC researchers, as Guckelsberger (2020, p. 78) points out:

- cognitive perspective: providing insights into the nature of creativity;

- engineering perspective: development of systems designed to foster human creativity, co-creative systems that take on some creative responsibility while interacting with others, and fully autonomous creative systems.

The focus of this thesis is related to the engineering perspective, exploring the nature of the interaction between user and system (from autonomous to co-creative) to foster human creativity. At this end of the continuum, researchers address the study and development of creative systems. Creative systems can be considered as fully autonomous systems or co-creative systems. A third category is identified by some authors (e.g. Karimi et al., 2018a), which corresponds to computational systems aimed at supporting and fostering human creativity.

Fully autonomous systems produce (creative) outputs whose creation is guided by an automatic evaluation process. While the evaluation may have human intervention (e.g. criteria weighting), the generation of artefacts is totally on the systems' side. Since the early days of CC as a research field, many fully autonomous creative systems have been proposed, using different approaches, such as evolutionary processes (e.g. Colton, 2012) or machine learning techniques (e.g. Heath and Ventura, 2016b).

In co-creative systems, there is a collaboration between humans and computers to produce creative artefacts. We will address co-creative systems in the following section.

3.1.3 *Computational Co-creativity*

Existing creative systems can be considered as being on a spectrum. On one end of the spectrum are located creativity support tools, which can help users in activities ranging from drawing or sketching (Dixon, Prasad, and Hammond, 2010; Lee, Zitnick, and Cohen, 2011) to typeface design (Cunha et al., 2016). These tools, however, are limited in terms of creative contributions to the process. On the other end are systems that exhibit autonomous creative behaviour – for example, some artificial artists are AARON (Cohen, 1988), *The Painting Fool* (Colton, 2011, 2012) or DARCI (Norton, Heath, and Ventura, 2013). These generative systems normally do not consider the feedback from the user.

Co-creative systems can be considered to be located somewhere between the two ends (Ravikumar, 2020, p. 3). These approaches to co-creativity establish a collaboration between several agents, one of which is required to be artificial. Research on human-computer co-creativity has been conducted with a greater focus on the human perspective, considering computers as tools rather than individual creators. This view is reflected, for example, in the classification of creative computational partners by Lubart (2005), who identifies four different roles for the computational agent: *nanny*, (the agent is responsible for routine tasks and manages user's time and work on creative tasks); *pen-pal* (the agent facilitates communication between human co-authors); *coach* (the agent advises the user, fostering their creativity); and *colleague* (the agent is creative on its own and able to contribute with ideas). Another categorisation is proposed by Maher (2012), who describes three different roles: *support* (the agent supports human creativity by providing tools and techniques), *enhance* (the agent extends the ability of the person to be creative) and *generate* (the agent generates creative ideas that the human interprets, evaluates or integrates as a creative product).

A key term in co-creativity is *collaboration*, which is defined by Terveen (1995) as "a process in which two or more agents work together to achieve shared goals". In this process, participants are normally considered equal, building on previous contributions and mutually influencing each other. Each brings its own knowledge and experience to the process, leading to different interpretations and unexpected solutions (Jordanous, 2017; Mamykina, Candy, and Edmonds, 2002). Ideas from different participants are put together, resulting in a scenario in which the sum is greater than the parts (Davis, 2013). As such, in a computational co-creative setup there is collaboration between human and computational agents in a shared creative process, in which agents contribute to the same goal, e.g. game level design (Yannakakis, Liapis, and Alexopoulos, 2014). Existing systems can be considered as hybrids between creativity support tools and generative systems, exploring different domains, such as:

- music: a percussion robot that mimics musicians and can generate synchronised improvisations (Hoffman and Weinberg, 2010);
- movement: an interactive art installation in which humans and artificial intelligent agents collaborate in movement improvisation (Long et al., 2017b);
- visual arts: a drawing system capable of contributing in a creative way by suggesting new elements (Davis et al., 2016).

Several characteristics should be present in a co-creative system. First, at least one of the agents taking part in the creative process needs to be artificial. Second, each agent should be able to perceive the contributions of other agents (Karimi et al., 2018a). Third, each agent should be able to contribute to the creative process, e.g. by expressing its own creative ideas, through autonomous action. As Karimi et al. (2018a) state, the degree of contribution of the agents does not need to be the same – different types of collaboration are accepted, e.g. *partnership* or *assistantship*.

Kantosalo and Jordanous (2020) analyse existing roles classifications for computational participants in co-creativity and propose a novel categorisation. In their categorisation, systems are firstly divided into *creativity support tools* and *co-creative colleagues*. Then, different categories are proposed (i-vii), based on the roles of the system:

- *Creativity Support Tools*: (i) support; (ii) train; and (iii) enable;
- *Co-creative Colleagues*: (iv) generate; (v) evaluate; (vi) find problems; and (vii) control initiative.

Another aspect concerning co-creative systems is related to ways of interaction. Rezwana and Maher (2021) present a framework that provides guidance to the exploration of the design space of interaction for co-creative systems. The framework is divided into *Components of Interaction between the Collaborators* and *Components of Interaction with the Shared Product*.

One last question worth mentioning is that despite the increasing scope of the capabilities of computational agents, they are not yet given the same agency as humans in co-creative contexts (Kantosalo and Jordanous, 2020). Nonetheless, co-creative systems are considered to foster the creativity of the user (Liapis et al., 2016) and recent developments focus on the assessment of their impact on ideation (Kim and Maher, 2021).

3.2 COMPUTATIONAL APPROACHES

There are many computational approaches that, in one way or another, involve the visual representation of concepts. These vary in terms of technique, degree of automation and kind of human-machine relationship. In this section, we provide an overview of the different types of approaches that concern the visual representation of concepts through computational means.

Different kinds of analysis could be used to organise the approaches related to the visual representation of concepts. First, a chronological order could be used to present the reader with a clear overview of the historical development of such systems. Despite being often useful, we believe that chronological ordering does not always provide the most interesting perspective. Second, an order could be established based on the kind of output produced by the approaches (e.g. abstract or pictorial; photorealistic or non-photorealistic). Although such framing is important for the goals of the thesis, we do not consider it as the main strategy when it comes to analysing related work. Third, approaches could be organised in terms of the nature of the interaction with the user (e.g. autonomous, creativity support tool or co-creative). However, the focus of this thesis is not human-machine interaction.

As such, we consider that the organisation of this chapter is a combination of all these types of analysis. The focus is on the degree to which concepts are represented in the system, going from systems in which there is a very low or absent encoding of conceptual knowledge, to those which have it as the main goal. Despite this, the reader will notice that the order in which the systems are presented is often chronological. In addition to the organisation in terms of conceptual representation, we grouped systems in terms of the type of system. Some systems may fit in more than one of these groups.

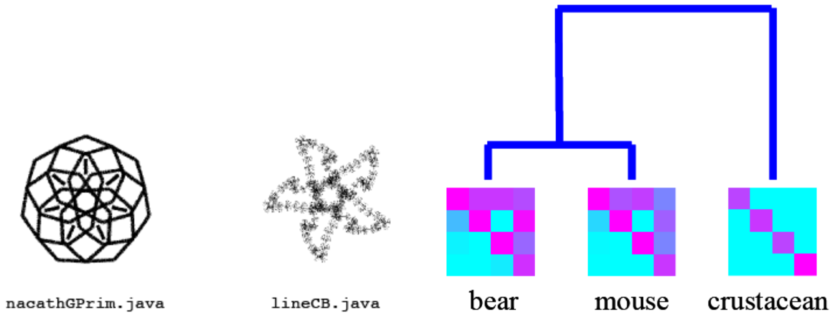
3.2.1 *Towards Pictorial Resemblance*

The appearance of computers enabled the creation of visual art through computational means. Earlier explorations involved creating black and white geometric drawings, e.g. the *Cubic Disarray* by Georg Nees (1968–1971), or the representation of digitised images, e.g. a version of *Mona Lisa* using pairs of decimal digits by Peterson (1965).

Beyond these initial explorations, authors started to work towards more pictorial approaches. An important example is AARON (Cohen, 1988), an AI system developed by Harold Cohen over several decades, going from an initial non-pictorial version to one that was able to use coloured pictorial representations. It is encoded with rules that represent Cohen's artistic concepts and its representations go from resembling children's drawings to, more recently, people in gardens. Similarly, ROSE (Representation Of Spatial Experience) was developed by Burton (1995) to model stages of children's drawings. These approaches are considered to be based on knowledge.

Other approaches of computational art involve the use of evolutionary approaches, in which evolutionary computation is employed in several artistic problems, as already mentioned in Section 3.1.1.2. The *Biomorphs* result from a famous example of an evolutionary system implemented by Dawkins et al. (1986), who demonstrated the artificial evolution of creature-like forms, guided by users.

Expression-based approaches were introduced by Sims (1991) to evolve images. These approaches were key for the development of systems that generate images that resemble certain objects. Machado, Correia, and Romero (2012) combine an expression-based evolutionary system with a Machine Learning (ML) face detector to automat-



(a) Visual identifiers (b) Intelligent icons

Figure 3.1: File identifiers: (*a*) Lewis et al. (2004) and (*b*) Keogh et al. (2006)

ically evolve face-resembling images. Similarly, Correia et al. (2013) use different ML classifiers to evolve faces, lips, breasts and leaves (see Fig. 3.2). Machado et al. (2015) combine object detectors to explore the automated evolution of ambiguous images. The goal in these images is that one can recognise two different objects, such as a face and a flower.

3.2.2 *Content Representation and Visualisation*

In addition to working towards the generation of images that resemble certain objects, other approaches focus on encoding data. A brief note should be made to the partial overlap between conceptual representation and data visualisation. As we will see in this thesis, conceptual representation can be used for visualisation purposes. We mention work that, despite being closer to data visualisation, has some interesting aspects that can be linked to the representation of concepts.

Figure 3.2: Evolved Lips. Source: (Correia et al., 2013).

The first example is the work by Lewis et al. (2004), who generate abstract visual identifiers for files using a shape grammar. Lewis et al. (2004) use a clustering algorithm based on the file names so that files with similar names will have similar identifiers. In their work, Lewis et al. (2004) disregard the actual content of the file, stating that their goal is not visualisation but rather to improve the recall of files by the user. In this way, we find ourselves at one end of representation: despite being based on data, these icons do not provide any information about it. In contrast, Keogh et al. (2006) propose the automatic generation of icons based on the contents of the files using a colour mapping approach, which they call *intelligent icons*.

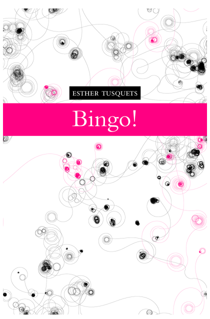
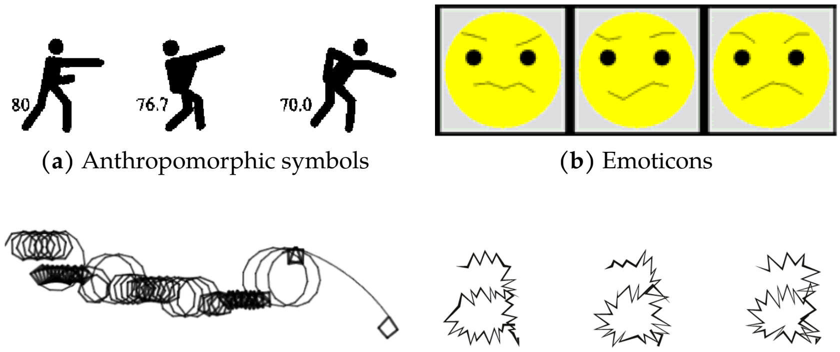


Figure 3.3: Cover for *Bingo!* by Esther Tusquets, produced by Data Book Covers by Pedro Cruz.

From the context of data visualisation, there are examples of the generation of visualisations based on the content of books. For example, the work *Data Book Covers*[3] by Pedro Cruz presents a tool that analyses the text of a book to identify the most frequent words, which can then be used to produce a visualisation based on the form of the text or con-

3 pmcruz.com/work/book-covers



(**a**) Anthropomorphic symbols
(**b**) Emoticons
(**c**) Lines
(**d**) Typeface characters

Figure 3.4: Different representations of "anger" and "rage": (*a*) Dorris et al. (2004), (*b*) Dozier et al. (2005), (*c*) Rodrigues, Cardoso, and Machado (2019) and (*d*) Maçãs, Palma, and Rebelo (2019).

tent of the book (see Fig. 3.3). Another example is the work by Duro, Machado, and Rebelo (2012), who also generate book covers based on the structure of the book, i.e. the number and length of chapters have an impact on the composition of the cover.

Another field that takes advantage of perceptual features for data representation is graphic design. As previously mentioned, there are several examples of dynamic visual identities that make use of visual characteristics of marks to encode data, e.g. the use of colour and shape in the graphic mark of the Nordkyn visual identity to represent live meteorological data.

3.2.3 *Representation of Emotions*

Dorris et al. (2004) use an IEC approach to evolve anthropomorphic symbols (Fig. 3.4a), in which the angles of nine limbs were represented by a vector of real-valued numbers. The goal was to evolve symbols that represent emotions. Similarly, Dozier et al. (2005) address emotion through the use of an interactive distributed evolutionary algorithm that allows the evolution of emoticons (Fig. 3.4b), whose mouth and eyebrows are line drawn based on vectors of integers. In these examples, the rotation and positioning of elements are used to produce certain body and face configurations, as a way to represent emotion.

A different approach is used by Rodrigues, Cardoso, and Machado (2019), who explore cross-modal associations between the domains of visual communication and perception of emotions by using a shape grammar with line visual features – such as *contour continuity*, *curve properties*, *stroke weight*, etc. – to produce images that depict different lines (Fig 3.4c), which aim to be perceived as related to certain emotions (*anger*, *calm*, *happy* and *sad*).

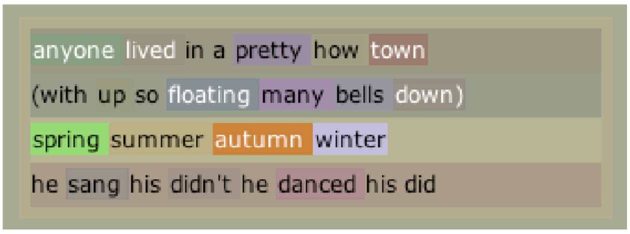


Figure 3.5: Colours assigned by *Colorizer* to part of E. E. Cummings poem. Source: (Havasi, Speer, and Holmgren, 2010).

Krcadinac et al. (2015) propose an abstract visualisation of emotions in chat messages using an animated particle simulation. They establish a mapping between emotion and image that combines *motion*, *colour* and *shape*, e.g. stronger emotions are represented with higher saturation and higher number of particles.

Maçãs, Palma, and Rebelo (2019) create a typeface that changes the *shape* of its glyphs based on emotional values retrieved from each sentence of a text (Fig 3.4d). This is done with a system called *typEm*, which receives a text as input, identifies emotions in it (*happiness*, *sadness*, *anger*, *fear*, *disgust*, and *surprise*) and outputs the texts written using the adapted glyphs.

Figure 3.6: #ESSYS* visual artwork. Source: (Seiça et al., 2021).

Seiça et al. (2021) present #ESSYS*, which is described as an online-based happening that represents emotions in *Twitter* data through the creation of environments that combine visual artworks and auditory compositions (Fig. 3.6). On the visual side, colour-emotion mappings are used based on the work by Plutchik (2001).

3.2.4 *Concept Representation through Colour and Shape*

The previous examples show how emotions can be represented through the use of *shape* and *colour*. Some authors go beyond emotions and explore *perceptual features* for the representation of concepts.

Havasi, Speer, and Holmgren (2010) present *Colorizer*, a program that takes into account physical descriptions of objects, as well as emotional connotations to guess a colour to represent a given input word or sentence (Fig. 3.5). To have access to knowledge on colour, they use data from: *ConceptNet*; *NodeBox*, which has a set of mappings between words and eleven pre-defined colours; and XKCD, which includes over 2.3 associations between colours and names. This approach generates single colours for single meanings.

Also concerning colour, Lin et al. (2013) present an algorithm that is able to map a set of categorical values (e.g. fruits) with a given target colour palette. They use *Google Image Search* to collect representative images for each value, which are then analysed in terms of colour distribution to determine the best assignment between values and colours.

Their approach is used in bar charts and said to lower response times, possibly due to reduced legend lookups.

A different approach to represent concepts is used by Long et al. (2017a). They develop an AI system that makes use of a symbolic visual language to represent words spoken by participants in a narrative context, resulting in a narrative art piece. The symbolic visual language is based on a set of pre-defined symbols (Fig. 3.7), which represent different kinds of agents (e.g. male agent, non-human agent, etc.), objects, prepositions and actions (reduced to a set of basic actions, e.g. *see*, *feel*, *be*, etc.). A set of colour to sentiment mappings is also used in the visual representation to fill the interiors of object and agent shapes, as well as a spiral that represents the overall sentiment of the story.

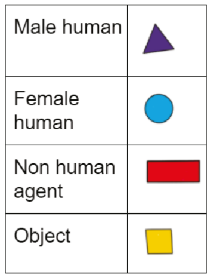


Figure 3.7: Symbolic visual language for narrative art piece by Long et al. (2017a).

3.2.5 *Visual Representation Support Tools*

Computational approaches have also been explored to assist users in the visual representation of concepts. These have addressed different goals, such as enabling the interaction among users, providing assistance to the representation of concepts, and fostering users' creativity through suggestions.

Piper (2010) uses a distributed approach, proposing an interactive genetic algorithm technique for designing safety warning symbols (e.g. *Hot Exhaust*). It uses symbol components as input which are then combined to produce new symbols. According to Piper (2010), this distributed approach allows the replacement of the usual focus group in symbol design process with a group of participants interacting using computers in a network.

Also on the subject of interaction among multiple users, Liapis et al. (2015) present a system called *Iconoscope*, which uses a game situation in which users play a guessing game to identify icons designed by other players for specific concepts.

Regarding support representation tasks, Zhao and Wang (2010) describe a system that helps the user to create icons. The user can draw their own icons and then the system is able to make recommendations to assist the creation process by presenting similar drawings or suggesting combination templates.

Garg, Berg, and Mueller (2011) propose a system to help users identify icons that can visually represent abstract concepts. The system has two main components: the *I-Bridge builder*, finds images and concepts related to a given input concept; and the *VIE-Designer*, which uses image clustering and iconisation to produce, for each cluster, an image that depicts common features of the objects in the cluster.

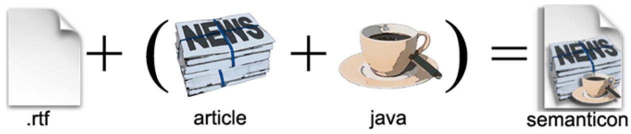


Figure 3.8: Production of a semanticon for "article_jave.rtf" by Setlur et al. (2005).

3.2.6 *Image Retrieval*

The approaches by Garg, Berg, and Mueller (2011) and Zhao and Wang (2010) lead us to a method commonly used in the visual representation of concepts, which involves the retrieval of images and their consequent use for representation. Going back to file icons, which we mentioned in the previous sections, several authors have explored the generation of icons with image retrieval and combination.

Setlur et al. (2005) describe a technique to automatically create multiple GUI icons, referred to as *semanticons* (Fig. 3.8), for a given file based on its name, path and contents. The data of the file is used to form a context, which is used to formulate a query to an image database to find corresponding images, based on keywords associated with each image. Then, a process is conducted to produce the icon, involving region extraction, image stylisation and composition (combining different retrieved images in a single icon through *juxtaposition*).

Roy et al. (2017) propose a method to assign a set of documents (files) with *topical icons*, which consists in (i) clustering the documents based on their content, (ii) followed by an extraction of significant phrases from documents of each cluster, which are then (iii) generalised using a concept taxonomy (*Wikipedia*'s taxonomy) and (iv) compared with icon tag lists (they use a set of 35,287 tagged icons from *Flaticon*[4] to (v) assign each document with an icon that represents its content. They report that their approach improves file navigation.

Image retrieval has been used for other purposes. In data visualisation, Setlur and Mackinlay (2014) propose the generation of semantically relevant icon encodings for categorical data values in scatter plots by using natural language processing and image retrieval. Featherstone (2019) proposes a strategy related to conceptual blending, in which the background and elements of a bar chart can be replaced using imagery related to the data thematic. The process involves the search for related topics using *ConceptNet* and *Wikipedia* pages, which are then used to gather clipart images – e.g. from an Iris Dataset, it gathers topics related to plants and eyes, which are then used to retrieve images. Coelho and Mueller (2020) propose an interactive support tool for designing *infomages* (infographics), which also uses image retrieval based on keywords from the data to be represented.

4 flaticon.com

Figure 3.9: Picture generated by the text-to-picture system by Zhu et al. (2007).

Figure 3.10: Artistic Typography by Tendulkar et al. (2019): replacing letters of "church" with thematic icons that visually resemble the replaced letter.

Related to graphic design, Tendulkar et al. (2019) propose an approach to the generation of artistic typography, which consists in the use of an unsupervised learning approach to build latent space that considers letters and cliparts (Fig. 3.10). This approach enables the composition of words that use icons (retrieved from the *Noun Project* based on a specified theme) that are visually similar to letters and are related to the topic of the word.

In music, Machida and Itoh (2011) propose the visual representation of pieces of music through the selection of icons based on musical features and lyrical keywords. For the auto-illustration of poems and songs, Schwarz, Berg, and Lensch (2016) present a pipeline that consists in collecting images from a large collection of annotated photos that match important words from the text and a specified style. To conduct this matching process, they compare image tags with word embeddings from the text, and perform a style prediction of the images. A related approach is followed in a supporting work of this thesis (Cruz et al., 2019). Differently, other authors have explored the composition of scenes that represent pieces of music, based on feature extraction (objects, locations, events, etc.) from lyrics and image retrieval (Kasai, 2013; Kikuchi and Kasai, 2012).

Some of the aforementioned approaches address the production of images from text using image retrieval which is commonly referred to as *text-to-image*.

3.2.7 *Text-to-image Approaches*

Text-to-image (or *text-to-picture*) involves the analysis of text and the selection of visual elements that can be used to represent the text (e.g. Goldberg et al., 2009; Ustalov, 2012). These approaches have been used for the illustration of instructions for patients (Bui et al., 2012), visual dialogue summarisation (Jiang, Liu, and Lu, 2016) and even audio-visual slideshow generation (Leake et al., 2020).

For example, Zhu et al. (2007) make use of the correlation between keywords and images and then search for the most likely image parts based on "picturable" text units to generate images (Fig. 3.9).

Inaba, Kanezaki, and Harada (2014) develop a system that automatically synthesises an object to an image when given the background image and the class name of the object. Their approach focuses on the learning of *spatial*, *scale-related*, and *appearance-based* contexts.

Mano, Yamane, and Harada (2016) describe a system to synthesise scene images from sentences. For this, they implement a hierarchical syntactic parser for sentence analysis, and explore the correlation between input sentences and image patches.

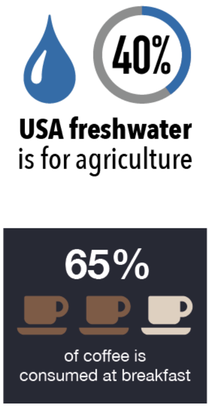


Figure 3.11: Examples generated by *Tex-to-Viz* for the statements "40% of USA freshwater is for agriculture" and "65% of coffee is consumed at breakfast". Source: (Cui et al., 2019).

Cui et al. (2019) propose a proof-of-concept system for *text-to-viz*, with the goal of automatically generating infographics from natural language simple statements, e.g. "More than 40% of USA freshwater is for agriculture" (see Fig. 3.11). The system is composed of two main modules: the *text analyzer*, which receives a textual statement from users and identifies segments (i.e. *entity types*, such as *modifier*, *whole*, *part* and *number*); and the *visual generator*, which generates a set of visual elements for different dimensions (*layout*, *description*, *graphic* and *colour*), based on the original statement and the identified segments, and synthesises infographic candidates that are shown to the user. To achieve this, they use a set of 100 common icons, which were manually assigned with keywords, and rely on pre-designed layout blueprints, which describe the different regions of the infographic.

The *text-to-image* systems described in this section are meant as illustrative examples. For a more detailed analysis of the research developed on *text-to-image*, we refer the reader to Hassani and Lee (2016) and Zakraoui, Saleh, and Al Ja'am (2019). Moreover, generative models have recently been employed to produce images from text input. While traditional *text-to-image* approachers lack the ability to generate new image content (only being able to change characteristics of the given images), generative models are able to produce new content. These will be addressed in Section 3.2.15.

3.2.8 *Collages*

Other approaches used for visual representation purposes involve *collage*. Collages can be defined as compositions made from the assemblage of different material forms (Huang, Zhang, and Zhang, 2011). Cook and Colton (2011) and Krzeczkowska et al. (2010) propose a collage-generation module for *The Painting Fool*[5] (Colton, 2011). The approach consists in extracting keywords from news articles from online sources (*Guardian News* website and *Google News* search, via their APIS), using them to retrieve images from *Google Images* and *Flickr* and

5 www.thepaintingfool.com

(**a**) War in Afghanistan

(**b**) Car

Figure 3.12: Examples of collages: (*a*) Cook and Colton (2011) and (*b*) collage by SOILIE (Breault et al., 2013)

then putting together image segments to generate *collages* (or *mashups*), which depict a particular news story (Fig. 3.12a).

Similarly, Breault et al. (2013) propose a system called SOILIE (Science of Imagination Laboratory Imagination Engine) that is based on a cognitive model of visual imagination – visual imagination is understood as generating mental simulations of world states (Fig. 3.12b). Their approach takes a single word and creates a static 2D scene that contains other elements related to the query word. To achieve this, they make use of large databases of labelled images, whose labels are used to identify spatial relationships between objects to be employed in the positioning of elements in the produced image. Further developments to the system are described in other publications, of which we highlight the work developed by Vertolli (2014).

Figure 3.13: "Fruit" Shin-chan collage. Source: (Huang, Zhang, and Zhang, 2011)

A different approach focuses on the generation of Arcimboldo-like collage (Huang, Zhang, and Zhang, 2011). The collage is produced from an input image, in which parts are replaced to produce a collage made from multiple thematically-related cutouts from retrieved images (the user is able to input descriptive keywords that are used in a text-based image searching). The output is described as representing the input image, disguised both in shape and colour, with the individual cutouts still being recognised (Fig. 3.13). To achieve this, Huang, Zhang, and Zhang (2011) employ techniques of cutout matching to make cutouts resemble structural components of the input image. In this case, the collage has a reduced degree of concept representation (only addressed by the common theme from the retrieved images).

Figure 3.14: Photorealistic Blend. Source: (Xiao and Linkola, 2015).

3.2.9 *Blending*

The previous collage-based approach leads us to one of the main topics of the thesis: *Blending*. When mentioning *Blending* there are two levels that need to be addressed – *conceptual* and *visual*.

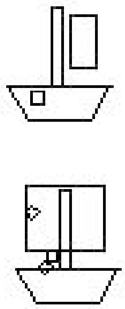


Figure 3.15: Boat-house blends by Pereira and Cardoso (2002)

Conceptual Blending (CB) theory is a cognitive framework proposed by Fauconnier and Turner (2002) as an attempt to explain the creation of meaning and insight. CB consists in integrating two or more *mental spaces* in order to produce a new one, the *blend(ed) space*. Here, *mental space* means a temporary knowledge structure created for the purpose of local understanding (Fauconnier, 1994).

On the other hand, Visual Blending (VB) consists in the production of new visual representations (e.g. *images*) by merging at least two existing ones (e.g. Fig. 3.14). In Computational Creativity (CC) research, VB can be combined with computational approaches to CB (Fauconnier and Turner, 2002) to produce representations for a blended mental space (e.g. Fig. 3.15). Some of the works that address blending are explicitly based on CB theory, as blending occurs at a conceptual level, while other approaches generate blends only at a representation or instance level by means of, for example, image processing techniques.

Current computational approaches to VB can be divided into two groups according to the *type of rendering* used: (i) *picture* or *photorealistic* rendering; and (ii) *non-photorealistic* (e.g. pictograms or icons). Examples of the first group are: the work by Steinbrück (2013), who combines image processing techniques with semantic knowledge gathering to produce images in which elements are replaced with similar-shaped ones (e.g. round medical tablets are transformed into globes); and *Vismantic* (Xiao and Linkola, 2015) – a semi-automatic system that produces visual compositions for specific meanings (e.g. *Electricity is green* is represented as the fusion between an electric light bulb and green leaves in Fig. 3.14). Examples of non-photorealistic blending approaches address, for example, the generation of pictograms or icons, e.g. the generation of visual representations for *boat-house* (Pereira and Cardoso, 2002) (Fig. 3.15) or the combination of signs (Confalonieri et al., 2015).

A categorisation can also be done in terms of where the blending process occurs: some interpret or visualise previously produced concep-

(a) Pictogram Montage

(b) Semantic hybridization

Figure 3.16: Blends focused on the visual side. Sources: (*a*) (Liu et al., 2016) and (*b*) (Aberman et al., 2018).

tual blends, e.g. Pereira and Cardoso (2002) experiment with conceptual blends produced for the input spaces *house* and *boat* (Fig. 3.15); others use blending only at the visual level, e.g. Correia et al. (2016) generate faces out of existing ones by recombining face parts (Fig. 3.17). A third approach, which can be called hybrid, consists in a process of blending that starts at the conceptual level and only ends at the visual level – we return to this topic in this thesis as we have developed work related to it, addressed in Chapter 6.

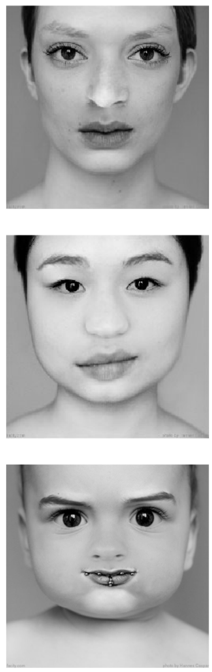
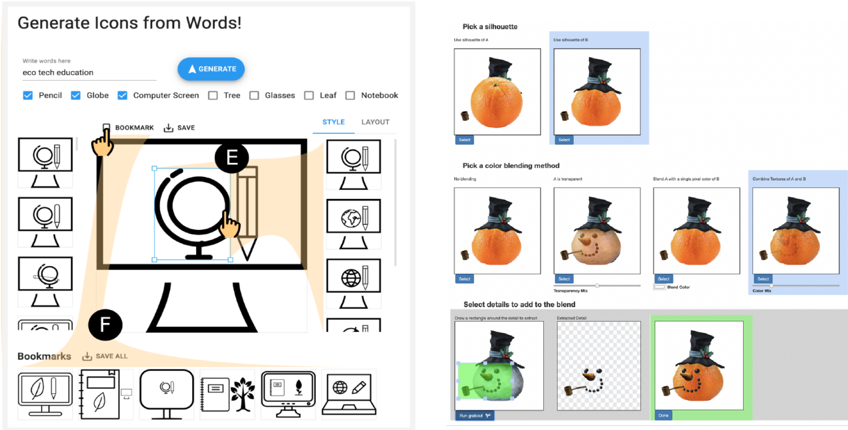


Figure 3.17: Faces generated by Correia et al. (2016) through combination of parts.

Starting with examples more focused on the visual side, Liu et al. (2016) present a system that produces new pictograms by combining parts of icons retrieved from a large online repository (Fig. 3.16a). They present four *icon customisation workflows*, two of which are particularly interesting for this thesis: *pictogram hybrids*, which consist in producing hybrids from stock pictograms; and *pictogram montage*, in which the system is guided by scribbles from the user that dictate the parts to maintain in the combination.

Aberman et al. (2018) present examples of what they call *semantic hybridization* (Fig. 3.16b), in which images are aligned based on correspondences identified using a Convolutional Neural Network (CNN). Then, based on neural activations, a low-resolution semantic mask is propagated to the original image, producing a hybrid image.

Chilton, Petridis, and Agrawala (2019) introduce a workflow for iterative design of visual blends, which they called *VisiBlends*. Users interact with the system to produce visual blends for a given concept, finding images that represent the concept and annotating them for shape and coverage. Then, the system automatically detects images that are suitable, according to shape similarity, and then automatically produces the blends. Chilton, Ozmen, and Ross (2020) combine automatic tools (grabcut extraction tool) with user interaction to iteratively improve the blends (*VisiFit* system in Fig. 3.18b).



(a) ICONATE interface (b) *VisiFit* blending steps

Figure 3.18: Systems for the production of blends. Sources: (*a*) (Zhao et al., 2020) and (*b*) (Chilton, Ozmen, and Ross, 2020).

Zhao et al. (2020) present ICONATE (Fig. 3.18a), which is a system that automates the process of generating icons given compound concepts and allows users to explore and customise the generated icons. To achieve this, the authors produced a dataset of 1000 compound icons annotated with semantic labels.

Sbai (2021) and Sbai, Couprie, and Aubry (2021) describe an approach for creating photorealistic blends from an image with a clear foreground object. The approach has two main tasks: (i) searching for images that have foreground objects that can be used as replacements; and (ii) conducting an automatic process of visual blending between the input image and the replacement one. In task (i), the goal is to find images with foregrounds that are visually similar yet semantically different. For this, the authors search images from different classes that have similar features to the query image.

On the other hand, approaches from the CC domain are often based on CB (Fauconnier and Turner, 2002) and give more importance to the conceptual side. The *boat-house* experience (Pereira and Cardoso, 2002) is, to the best of our knowledge, one of the earliest attempts to computationally produce visual blends with a conceptual grounding. The work was motivated by the need to interpret and visualise blends produced by a preliminary version of the *Divago* framework, which is one of the first artificial creative systems based on CB theory (Pereira, 2007). In addition to a declarative description of the concepts via rules and concept maps (i.e. graphs representing binary relations between concepts), Pereira and Cardoso (2002) also considered a domain of instances, which were drawn using a Logo-like programming language (Fig. 3.15). To test the system, the authors performed several experiments with the *house* and *boat* blend (Goguen, 1999), considering different instances for the input spaces.

Ribeiro et al. (2003) explore the use of the *Divago* framework in procedural content generation. In this work, the role of *Divago* was to produce novel creatures at a conceptual level from a set of existing ones. Then, a 3D interpreter was used to visualise the objects. The interpreter was able to convert concept maps from *Divago*, representing creatures, into *Wavefront* OBJ files that could be rendered afterwards (Fig. 3.19).

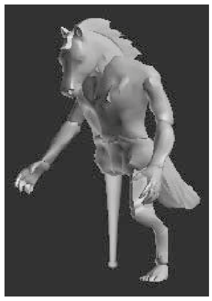


Figure 3.19: Creature produced by Ribeiro et al. (2003).

Steinbrück (2013) introduces a framework that formalises the process of CB while applying it to the visual domain. The framework is composed of five modules that combine image processing techniques with gathering semantic knowledge about the concept depicted in an image with the help of ontologies. Elements of the image are replaced with other unexpected elements of similar shape (for example, round medical tablets are replaced with pictures of a globe).

Confalonieri et al. (2015) propose a discursive approach to evaluate the quality of blends. The main idea is to use Lakatosian argumentative dialogue (Lakatos, 1976) to iteratively construct valuable and novel blends as opposed to a strictly combinatorial approach. To exemplify the argumentative approach, the authors focused on icon design by introducing a semiotic system for modelling computer icons. Since icons can be considered as a combination of signs that can convey multiple intended meanings to the icon, Confalonieri et al. (2015) propose argumentation to evaluate and refine the quality of the icons.

Figure 3.20: "Electricity is green" blend by Xiao and Linkola (2015) (rotated).

Xiao and Linkola (2015) proposed *Vismantic*, a semi-automatic system aimed at producing visual compositions to express specific meanings, namely abstract concepts. Their system is based on three binary image operations (*juxtaposition*, *replacement* and *fusion*), which are the basic operations to represent *visual metaphors* (Phillips and McQuarrie, 2004). For example, *Vismantic* represents the slogan *Electricity is green* as an image of an electric light bulb in which the wire filament and screw base are fused with an image of green leaves (Fig. 3.20). The selection of images as well as the application of the visual operations require the user's intervention.

In terms of character blending, one example is the blend of *Pokémon* by Onsager (2013),[6] who produces a blended image and name for each *Pokémon* pair (Fig. 3.21).

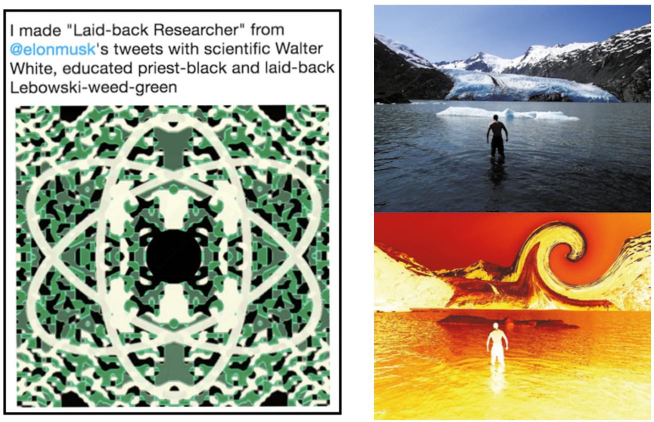
3.2.10 *Blending through Perceptual Transformation*

Another work on the topic of *Pokémon* is presented by Liapis (2018), who produces mappings between type and attributes (such as *colour*, *shape* and *in-game* sprite), which allow the change of type of a *Pokémon* – a type change to *fire-ground* modifies the colour to reddish tones (see Fig. 3.22). This example by Liapis (2018) introduces a different kind of blending: a given image is changed by combining it with a certain perceptual aspect, for example, a different colour.

Figure 3.22: Changing *Pokémon* type by Liapis (2018) (adapted image).

6 pokemon.alexonsager.net

Figure 3.21: Blending of *Pokémon* image and name by Onsager (2013).



(**a**) ElonMusk (**b**) "fiery"

Figure 3.23: Perceptual Transformation Blends: (*a*) Metaphor based on the affective profile of ElonMusk by (Veale, 2018); and (*b*) image rendered by DARCI from a source image and based on the concept "fiery" (Heath and Ventura, 2016b).

Figure 3.24: Lemotif: visualise salient themes and emotions from text (Li and Parikh, 2020).

Veale and Cook (2018) describe the development of a *Twitter* bot that produces tweets based on the analysis of celebrities' *Twitter* posts. The tweets result in metaphors based on the affective profile of the person accompanied by a visual metaphor, grounded in linguist content (Veale, 2018). The images are a combination of abstract and figurative elements, meant to capture the tweet content, making use of semantically related emoji and colours (see example[7] in Fig. 3.23a).

Heath and Ventura (2016b) apply an approach called *Associative Conceptual Imagination framework* to learn associations between low-level image features and adjective vectors from Vector Space Models, using a large neural network. This approach is used in a system called DARCI (Heath, Norton, and Ventura, 2014; Norton, Heath, and Ventura, 2011), which renders images that represent a given concept (e.g. an image rendered based on the concept "fiery" in Fig. 3.23b).

Li and Parikh (2020) present a system that uses machine learning to analyse text-based input journal entries and generate representations of detected themes and emotions in the form of image motifs. It uses a set of icons related to different thematics and colours related to various *feelings* or *emotions* (e.g. yellow is linked to "happy" in Fig. 3.24).

7 twitter.com/BotOnBotAction/status/946411535647301633

(a) (b)

Figure 3.25: Neural Network generated images: (*a*) image generated for the adjective "fiery" with a DNN trained on the DARCI dataset (Heath and Ventura, 2016a); (*n*) image generated by applying deep dream on a white-noise image (Berov and Kuhnberger, 2016).

3.2.11 *Connectionist Approaches*

Some of the already described work uses symbolic approaches, while others explore connectionist-based ones, such as artificial neural networks. As an example, the work with DARCI by Heath, Norton, and Ventura (2014) and Norton, Heath, and Ventura (2011) (already mentioned) employs artificial neural networks to automatically render images that match a given set of adjectives. For this, they use multiple neural networks, which correspond to specific adjectives and model how humans identify the adjective in images. As mentioned by Ventura (2019), the development DARCI is centred on the importance of communicating meaning in art, resulting in the creation of images that intentionally express a given concept using visual metaphor.

Mordvintsev, Olah, and Tyka (2015) introduce *DeepDreams*, which produces hallucinatory-like imagery by using a CNN architecture. The approach can be considered as one of the important steps towards the use of Deep Neural Networks (DNNs) for the generation of visual art, leading to the appearance of several deep learning methods for image generation (e.g. Heath and Ventura, 2016a) (Fig. 3.25a).

Despite not always being focused on the representation of concepts, these approaches can be used to produce images in which one may identify some pictorial features. An example is the work by Berov and Kuhnberger (2016), who propose a computational model of visual hallucination based on deep neural networks (Fig. 3.25b). To some extent, the creations of this system can be seen as visual blends.

Nguyen et al. (2016) introduce a deep generator network (DGN), which produces images that maximise the activation of certain neurons of the neural network, thus working as a visualisation of the features that trigger the neuron.

Other approaches, such as *DeepStyle* by Gatys, Ecker, and Bethge (2015), can also be seen as a form of visual blending. *DeepStyle* is based on a deep neural network that has the ability to separate image content from certain aspects of style, allowing to recombine the content of an

Figure 3.26: Class-conditional sample generated by *BigGAN* (Brock, Donahue, and Simonyan, 2018).

Figure 3.27: Averaged pictures of object categories: mountain (Torralba and Oliva, 2003).

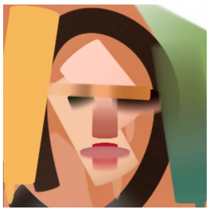


Figure 3.28: Abstract Face. Source: (Mellor et al., 2019)

Figure 3.29: Hidden Birds by Matty Mariansky

arbitrary image with a given rendering style – referred to as neural style transfer. The system is known for mimicking features of different painting styles. Several other authors have seen the potential of deep neural networks for tasks related to visual blending (e.g. McCaig, DiPaola, and Gabora, 2016).

Another architecture is known as Generative Adversarial Network (GAN) (Goodfellow et al., 2014) and consists in the use of two models, a generator and a discriminator, which are trained simultaneously with opposite objectives, competing with each other. The discriminator is a classification network that has the role of assessing whether images are real or fake/generated. On the other hand, the generator is used to produce new images and is optimised with the goal of fooling the discriminator. GANs are often used for tasks of *image-to-image translation* (e.g. Sun et al., 2019) and *text-to-image generation* (e.g. Reed et al., 2016), being widely used for generative purposes by computational artists (Machado, 2018) – e.g. GAN-generated pieces from Anna Ridler and Mario Klingemann were recently auctioned at *Sothebys* auction house.[8] Sbai (2021) highlights directions that have shaped the research on GANs, identifying challenges such as the development of techniques that allow better convergence guarantees, the development of techniques that enable the generation of high-resolution images and the proposal of metrics for the evaluation of generated images. Moreover, while some GANs have been trained for the generation of specific kinds of images, others are able to generate images of multiple classes – e.g. *BigGAN* (Brock, Donahue, and Simonyan, 2018), see Fig. 3.26. Different types of GAN architectures have been used for several purposes, such as the generation of emoji (Radpour and Bheda, 2017), icons (Yang et al., 2021) and logos (Jain et al., 2021). Four main existing approaches to image editing with generative models are identified by Sbai (2021): image to image translation, style transfer, image manipulation in a latent space and attribute based image editing.

3.2.12 *Generating Concept Abstractions*

Some approaches address what can be considered as abstractions of concepts. These abstractions can go from averaged pictures of object categories (Torralba and Oliva, 2003) (Fig. 3.27) to more elaborated compositions that retain aspects that trigger the observer into thinking about a given concept.

Some authors focus on the abstraction of specific objects. Mellor et al. (2019) describe an approach that uses reinforcement learning agents that are trained with rewards provided by a discriminator network to produce abstract faces using brush strokes (see Fig. 3.28). The work *Hidden Birds*[9] by Matty Mariansky (see Fig. 3.29) consists of a bird gen-

8 sothebys.com/en/buy/auction/2021/natively-digital-a-curated-nft-sale-2
9 supersize.co.il/portfolio/birds/

(**a**) Tick (**b**) Bee

Figure 3.30: Concept Abstractions: (*a*) abstraction of "Tick" (2018) (White, 2019); and (*b*) neural painting of *optimal* bee (Nakano, 2019)

erator that combines a differentiable rasteriser for vector graphics (Li et al., 2020) with a CLIP model (Radford et al., 2021) trained on text-image pairs (more detail on CLIP is given in Section 3.2.15), to assess the resemblance of the generated images to birds.

With a wider conceptual reach, the work *Perception Engines* by White (2019) obtains similar abstract results by using convolutional neural networks to generate images for classes of ImageNet, which are composed of lines of different thickness and colours (Fig. 3.30a).

Another approach to abstract representations is the work by Nakano (2019), who describes the implementation of a neural painter that optimises brushstrokes to activate neurons in a pre-trained convolutional network, thus generating what the authors refer to as "ideal" paintings of each class (Fig. 3.30b).

3.2.13 *Sketch Representations*

Another type of generalisation involves sketching. The *Google* web tool *Quick, Draw!* (Jongejan et al., 2016) allows the user to draw and would try to guess the object being drawn. Ha and Eck (2017) use sketches drawn on *Quick, Draw!* to train *Sketch-RNN*, a recurrent neural network capable of generalising concepts (e.g. *cat*) in order to draw them. Moreover, it is able to generate interpolations between several concepts (e.g. a *pig*, a *rabbit*, a *crab* and a *face*). This can be used as a tool for representing new concepts through visual blending – e.g. generating sketches of *chair* trained on *cat* drawings results in cat-styled chairs (Fig. 3.31). These examples mainly focus on visual representation and neglect conceptual aspects.

Other approaches use machine learning techniques to teach systems to draw based on user-drawn sketches. One example is the co-creative system *Drawing Apprentice* (Davis et al., 2016), which uses convolutional neural networks to perform real-time object recognition on the

Figure 3.31: Latent space of generated cats conditioned on sketch drawings of chairs by Ha and Eck (2017).

user sketch and responds with drawings of related objects. Another example is the work by Fan, Dinculescu, and Ha (2019), who present a web-based system for collaborative sketching, in which a recurrent neural network is used to make an artificial agent collaborate with a human agent.

Karimi et al. (2018b) describe a computational model of conceptual shifts based on sketches, using a dataset of *Quick, Draw!* sketches (Jongejan et al., 2016). The approach employs a CNN-LSTM model and uses sketch similarity to find sketches that are visually similar to a given sketch yet belonging to different categories. The authors introduce the possibility of using the approach to aid humans in blend production.

Mihai and Hare (2021) introduce a system with the goal of establishing emergent visual communication between agents, which communicate through stroke drawing. The agents are parameterised by deep neural networks and trained using standard gradient techniques.

A different approach is used by Huang and Canny (2019) and Huang et al. (2020), who present a machine-learning-driven system that allows users to interact through text instructions to produce a sketched scene. The system generates scenes by sketching objects based on the specifications given by the user. This last example is also within what can be referred to as *text-to-image* synthesis using neural networks, which we will further address in Section 3.2.15.

3.2.14 *Neural Network Blend Generators*

As mentioned in regards to the work by Ha and Eck (2017), neural networks can be explored for visual blending purposes. While some approaches focus on improving the task of seamlessly blending an object from a source image onto a target image, e.g. combining the *Poisson* image blending method with neural networks (Wu et al., 2019;

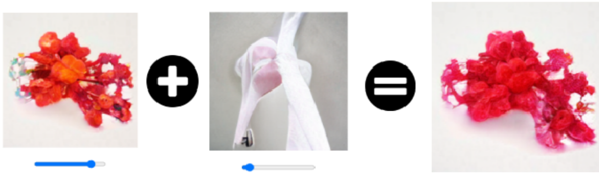


Figure 3.32: Mechanics of *crea.blender* (Rafner et al., 2020).

Zhang, Wen, and Shi, 2020), others give more attention to the generation of images that can be considered blends, e.g. generating cat-styled chairs (Ha and Eck, 2017). An interesting example is the work by Matty Mariansky involving *StyleGAN* halfway training between the *Flickr-Faces-HQ* dataset, which contains portraits of people, and a beetle dataset[10] – resulting in the generation of blends between faces and beetles (Fig. 3.33).

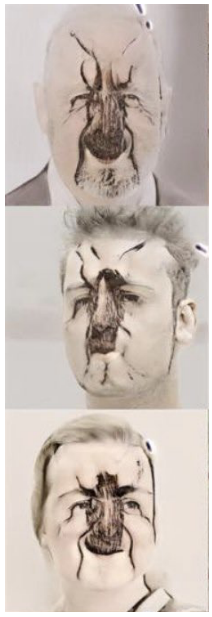


Figure 3.33: Blends between faces and beetles using *StyleGAN* by Matty Mariansky.

Chen (2019) and Chen et al. (2019a) present a GAN model that receives two concepts and is able to produce blended images that reflect features of both concepts (e.g. a *spoon* and a *leaf*). In this case, two discriminators are used, one for each concept, and the generator's task is to produce images that cheat both discriminators.

Another example related to visual blending is *crea.blender* (Rafner et al., 2020), a GAN based system that blends existing images (Fig. 3.32). Similarly, the platform *Art Breeder*[11] allows users to remix images through the modification and combination of latent vectors.

Bau et al. (2020) propose the modification of images by manipulating the intermediate layers in GAN models. They describe how the change of rules can enable the user to establish a copy/paste procedure based on an element of a given image (e.g. a hat or glasses), which may result in generating images in which, for example, horses have hats (Fig. 3.34).

Other approaches that can be considered as related to blending address the generation of images based on text, for example through the use of language models to guide neural networks (e.g. Ge and Parikh, 2021). These will be addressed in the next section.

Figure 3.34: Horse with hat through GAN layer manipulation (Bau et al., 2020).

3.2.15 *Text-to-Image Using Neural Networks*

Approaches that focus on *text-to-image* generation are particularly interesting to us, having as main goal the visual representation of concepts. Mei, Zhang, and Yao (2020) refer to the topic as "language to vision" and identifies two key challenges: (i) the interpretation of language input and (ii) the alignment between visual and textual modalities. For example, Mansimov et al. (2015) uses a variational recurrent autoencoder with attention to generate images from text captions.

10 supersize.co.il/portfolio/beetlegan/
11 artbreeder.com/browse

Other authors explore the capabilities of GAN architectures. Reed et al. (2016) experiment with a GAN approach to produce images from visual descriptions of, for example, flowers and birds. Puyat (2017) produces emoji from a search query, using a pre-trained *word2vec* model to convert text into an embedding that can be used by a GAN. Similarly, Chen et al. (2019b) present *MemeFaceGenerator*, which generates meme-faces from text inputs. Mittal et al. (2020) explore multi-modal input by producing images from an incomplete sketch and a handwritten word.

Figure 3.35: Portrait edit through latent optimisation (top image is the input and the bottom ones are obtained using the corresponding text promp) (Patashnik et al., 2021).

Hong et al. (2019) propose a hierarchical approach to *text-to-image* generation by firstly constructing a semantic layout from the input text, which is then converted to an image. For more information on the use of different GAN architectures for *text-to-image* purposes, we refer the reader to Agnese et al. (2020) and Frolov et al. (2021). Agnese et al. (2020) survey existing approaches of *text-to-image* synthesis using GANs, proposing a taxonomy of four major categories: *semantic enhancement*, *resolution enhancement*, *diversity enhancement* and *motion enhancement*.

Radford et al. (2021) present the *Contrastive Language-Image Pre-training* (CLIP) model, which was trained on 400 million text-image pairs and can be used to estimate the semantic similarity between a given text and an image – the cosine distance between encodings is smaller the more related an image-text pair is. The technique employed is called *contrastive learning* and is used to compress two models at once, resulting in similar latent representations of two different, yet related, media items (Colton et al., 2021). Several authors have explored CLIP for *text-to-image* generation (e.g. Galatolo, Cimino, and Vaglini, 2021).

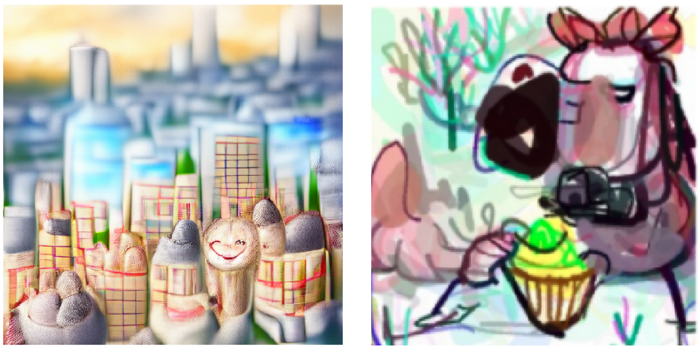
Patashnik et al. (2021) explore CLIP models in combination with *StyleGAN* image manipulation, which enables the manipulation of input images through text prompts (Fig. 3.35).

Galanos, Liapis, and Yannakakis (2021) present *AffectGAN*, which generates images based on affect using two models (Fig. 3.36a): CLIP (Radford et al., 2021) is used to include semantic information on the generation process; and a convolutional VQGAN model, which has a hybrid architecture with both transformer and GAN elements.

Frans, Soros, and Witkowski (2021) present *CLIPDraw*, an algorithm that uses CLIP as a metric to maximise the similarity between a text description and a generated drawing based on vector strokes (Fig. 3.36b). Schaldenbrand, Liu, and Oh (2021) combine *CLIPDraw* with style loss to develop *StyleCLIPDraw*, allowing the control of style of the synthesised drawings.

Ramesh et al. (2021) describe DALL·E,[12] which is a transformer-based text-to-image zero-shot pre-trained model with 12 billion parameters. Their approach demonstrates the potential for *text-to-image* generation of multi-modal pre-trained models (Fig. 3.37). Of special interest are

12 openai.com/blog/dall-e/



(**a**) "A happy cityscape" (**b**) "Horse eating a cupcake"

Figure 3.36: Images generated resorting to CLIP: (*a*) (Galanos, Liapis, and Yannakakis, 2021); and (*b*) (Frans, Soros, and Witkowski, 2021)

the results for concept combination, e.g. "an armchair in the shape of an avocado". A similar work by Ding et al. (2021) reports better results than DALL·E, by introducing a *Vector Quantised Variational AutoEncoder* (VQ-VAE) tokeniser. Other examples exist, such as *Emojich*,[13] a model with 1.3 billion parameters trained with text-emoji pairs based on the multi-modality big pre-trained transformer *ruDALL-E Malevich* (a Russian recreation of DALL·E model), which is able to generate emoji-style images (Fig. 3.38).

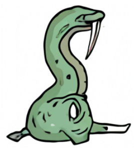


Figure 3.37: Image generated with DALL·E for the prompt "a professional high quality illustration of a snake walrus chimera. a snake imitating a walrus. a snake made of walrus".

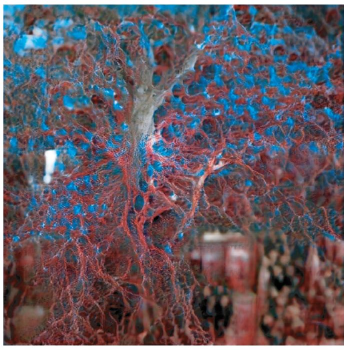
Figure 3.38: Image generated with *Emojich* for the prompt "a wolf in sheep's clothing".

Colton et al. (2021) introduce the notion of generative search engine, in which images are produced instead of retrieved (Fig. 3.39a). For this, they explore *text-to-image* generation using the implementation of Big Sleep,[14] in which CLIP (Radford et al., 2021) guides the *BigGAN* generator.

Ge and Parikh (2021) use large-scale language and image generation models to generate visual blends (Fig. 3.39b). In their approach, only one object is given as input (e.g. *moon*) and then a relevant object is identified by using prompt-engineering with language models (*reasoning phase*). Two strategies are described to identify relevant objects: *simile-inducing* and *property-guided*. In simile-inducing, a prompt like "the moon is like a [MASK]" leads to relevant objects, e.g. *ghost*. In property-guided, a specific property, e.g. *shape*, is used in a prompt like "The shape of the moon is [MASK1]´´ predicting the shape of the input concept as "spherical". Then a second prompt is used to identify relevant objects, e.g. "The shape of the [MASK2] is spherical" identifies *shell*. After identifying a relevant object (e.g. *ghost* in the *simile-inducing* strategy), a property of the object is retrieved using a prompt like "the ghost has the property of [MASK]", which leads to the property *dead*. Finally, a description of the blend is produced based on the input object, the relevant object and its property, e.g. "a moon that is dead like a

13 huggingface.co/sberbank-ai/rudalle-Emojich
14 github.com/lucidrains/big-sleep

(**a**) "A skyline in the style of Mondrian"



(**b**) "A tree made of blue and red blood vessels"

Figure 3.39: CLIP-guided GAN Image generation: (*a*) (Colton et al., 2021); and (*b*) (Ge and Parikh, 2021)

ghost". Then in the *visualisation phase*, an image is generated based on the produced description, using text-based image generation models – *BigSleep*, *DeepDaze*[15] and DF-GAN (Tao et al., 2020).

3.3 SUMMARY

In this chapter, we have introduced the reader to concepts that are core to this thesis. We started by making a brief introduction to *Computation Design*, *Computational Creativity* and *Computational Co-Creativity*. Then, we described existing computational approaches that are related to the visual representation of concepts. The presented approaches vary in terms of conceptual reach, going from approaches that produce images in which there is only some pictorial resemblance without a clear intention to represent concepts, to approaches that address *text-to-image* generation. Moreover, we presented approaches of different types of output, e.g. *icons*, *collages*, *blends*, etc. Even though the focus of this thesis is on symbolic approaches, we have also described work that uses connectionist ones. As a last remark, we would like to mention that the connection between *symbolic* and *connectionist* strategies may also be explored in the future for the visual representation of concepts (Aggarwal and Parikh, 2020).

15 github.com/lucidrains/deep-daze

Part II

TURN ON THE BLENDER

It all has to start somewhere: the reading of the first page, the writing of the first words, the implementation of the first function or, as we prefer to see it, the click of a blender's ON button.

In this part, we describe the start of our explorations, which go from the analysis of blends with the goal of gathering knowledge, to the implementation of a system that connects the conceptual side with the visual side.

4 BLENDING FOR CONCEPT REPRESENTATION

Indeed, blends seem to be omnipresent in everyday life,
filling up nearly every corner of our existence,
regardless of our awareness or unawareness of their presence.

— Adam T. Warchoł (Warchoł, 2018, p. 75)

In the previous chapter, we described existing computational approaches that are related to the visual representation of concepts. Some of the approaches are based on a kind of creativity that is referred to as *combinational* by Boden (1998), in which preexisting ideas and concepts are combined in ways considered novel. Related to this is the notion of *bisociation* ("bisociation of matrices") introduced by Koestler (1964). Koestler (1964) proposes that creativity occurs based on the unusual overlap of ideas drawn from different domains of knowledge (see Fig. 4.1). This combinational view has been studied by Fauconnier and Turner (2002), who proposed a framework for conceptual blending in which two input concepts are combined to produce a new one called the blend.

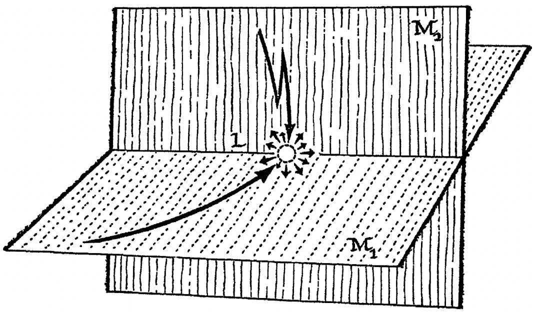


Figure 4.1: Illustration of Koestler's theory of bisociation. The situation or idea L, in which two self-consistent but normally incompatible frames intersect (M_1 and M_2). Source: (Koestler, 1964).

On the other hand, as explored in Chapter 2, visual language systems often use combinatorial strategies, producing new symbols by combining existing ones. These strategies are especially useful for concepts that do not have a simple translation into a visual representation, as pointed out in Section 2.2.2.

For example, consider the concept *English Pub*. As Ungerer and Schmid (2006, p. 39) point out, it is difficult to picture a line drawing of *English Pub*. However, a possible solution may resort to the combination of *house* or *building* and *beer*, which may be used for its representation, despite not covering the full scope of what an *English pub* is. Another example is the concept *mammal*, which comprises many different animals, e.g. dogs, mice, elephants, etc. In this sense, there is not a common shape that can be used to represent *mammal* but, as Ungerer and Schmid (2006, p. 86) state, the concept highlights important attributes such as "born from the mother's womb" and "fed by milk from the mother's body". Following this viewpoint, a combination between different visual representations (e.g. baby, mother and milk) may provide a way to represent the concept *mammal*.

For the visual representation of concepts, we identify two levels in which combinatorial processes are used: *conceptual* and *visual*.

In this chapter, we will address these two levels by first briefly introducing the reader to *conceptual combination* and *blending*, and then focusing on *visual blending*.

4.1 CONCEPTUAL COMBINATION AND BLENDING

In section 2.2, we provided an introduction to *concepts*, which we defined as dynamic abstractions that refer to ideas, objects or actions. When it comes to producing novel concepts, one can recall different tasks, such as *conceptual invention* and *formation*, which involve the discovery and creation of novel concepts. However, the creation of concepts is not the focus of this thesis. Instead, we are interested in how existing concepts can be used to (visually) represent other concepts through combination.

Briefly, the process of combining two or more concepts into a novel concept is referred to as *conceptual combination*. Different cognitive processes can be been identified as having a role in how concepts are combined, among which conceptual blending (Chan et al., 2017). Overall, conceptual combination can be considered to encompass two main topics of study: one related to *linguistics*, involving the process of how words can be put together to create novel terms; and one related to how concepts are combined in *cognition*, giving rise to novel concepts. The two levels are interrelated, although sometimes studied separately.

4.1.1 *Combining concepts*

On the linguistic level, the study of conceptual combination often concerns language compositionality and the interpretation of compounds. For example, in adjective-noun compounds, e.g. *red ball*, the nominal part ("ball") can be seen as the basic level category, which is modified by the adjectival element ("red"), interpreted as the assignment of a colour to the ball. In this way, a modifier-head structure exists, in which the first element modifies the second (Ungerer and Schmid, 2006).

Differently from adjective-noun compounds, in noun-noun combinations, e.g. *apple juice*, there are two basic level nominal categories (e.g. *apple* and *juice*). Despite using the same word formation, noun-noun compounds have a more complex way of functioning.

Costello and Keane (2000, 2001) propose different types of interpretation for noun-noun compounds, mentioning that some are more common than others and that other types may exist. The five types proposed by Costello and Keane (2000, 2001) are:

- Relational – a relation is established between the two parts (e.g. an *apartment dog* is a small dog that lives in an apartment);
- Property – a new concept is created by transferring a property from one concept to the other (e.g. an *elephant fish* is a big fish);
- Conjunctive – the combined concept is an instance of both parts (e.g. *pet bird* is both a bird and a pet);
- Hybrid – the combined concept is a blend of the two concepts (e.g. *drill screwdriver* is a two-in-one tool with features of both a drill and a screwdriver);
- Known-concept – the compound describes another concept related to the ones being combined (e.g. *cow house* describes *byre*).

From the examples given in these types, one may notice how the noun-noun compounds can also be seen as having a *modifier element* (the first word) and a *head element* (the second word). According to Costello and Keane (2001), properties are normally transferred from the *modifier* concept to the *head*.

However, this modifier-head structure does not always work in the same way – the second element of the compound is not necessarily the conceptually dominant part (Ungerer and Schmid, 2006). For example, in *orange juice* one may extract attributes like "liquid" and "served in glasses" from the second element ("juice"), and "made from oranges" and "orange in colour" from the first element ("orange"), which has a high impact on the meaning of the compound. This shows that the first element can be seen as having equal or even greater importance than the second one. Moreover, different kinds noun-noun compounds exist. For example, *orange juice* is based on *type-of* relationship and *shoelace* is based on *part-whole*.

Costello and Keane (2000) describe what is referred to as the Constraints theory, which focuses on explaining the creativity and efficiency of conceptual combination. According to Costello and Keane (2000), in a process of conceptual combination three constraints should be satisfied: *diagnosticity*, *plausibility* and *informativeness*. Diagnosticity gives importance to the presence of diagnostic properties from each of the elements of the compound, for example "prickly" is more diagnostic of *cactus* than "green", hence a *cactus fish* is more likely to be a *prickly fish* than a *green fish*. Plausibility gives importance to elements that have occurred together, based on past experience. For example, an *angel pig* as a "pig with wings on its torso" is more plausible than as a "pig with wings on its head". Informativeness gives importance to interpretations that communicate something new, e.g. an interpretation of *pencil bed* as a "bed made of wood" would be low on informativeness, whereas a "bed with a pencil-like shape" would be high. Other theories of conceptual combination exist, e.g. *Composite Prototype* (Hampton, 1987), but they are not within the scope of this thesis.

An aspect to be considered in compounds is *compositionality*, which can be seen as the degree to which the meaning of a compound can be derived from its parts (Roberts and Egg, 2018). For example, *apple juice* is a compound with a high degree of compositionality.

Other compounds have a lower degree of compositionality, being strongly linked to aspects that are not directly related to the elements that constitute them. An example is *wheelchair* (Ungerer and Schmid, 2006), which does not solely rely on the categories *wheel* and *chair* but also draws from others, such as *invalid*. The meaning derived from these additional categories can be seen as an emergent conceptual structure, which goes beyond the process of conceptual combination.

4.1.2 *Conceptual Blending*

Fauconnier and Turner (2002) introduce *Conceptual Blending* (or *Conceptual Integration*) as a cognitive operation in which two *input mental spaces* are brought together, producing a new *blended space*. Their work was initially motivated by the study of cognitive phenomena, such as *Metaphor* and *Analogy*, which they identified as only a subset of the range of conceptual blending phenomena (Fauconnier and Turner, 1998). Mental spaces, which are key in the definition of Conceptual Blending, are understood as conceptual packets that are dynamically built during discourse (Fauconnier, 1994), composed of elements (concepts) and connections among them, being context-dependent and not necessarily faithful representations of reality (Ungerer and Schmid, 2006; Warchoł, 2018). With their proposal of Conceptual Blending, Fauconnier and Turner (2002) provide a notion of emergence of meaning (as observed in *wheelchair*), based on the assumption that language and thought are not strictly compositional and do not rely exclusively on projection processes (Evans and Green, 2006, p. 402) – i.e. a blend is considered to be more than the sum of its component parts.

In Conceptual Blending, two (or more) *input spaces* are linked by using a third space called *generic space* (see Fig. 4.2). The generic space has information that is common to the input spaces, allowing mapping to occur between its elements and their counterparts in each of the input spaces. The mapping then motivates the establishment of *cross-space mappings* between the input spaces. When creating these connections between spaces, different kinds of integration networks may be

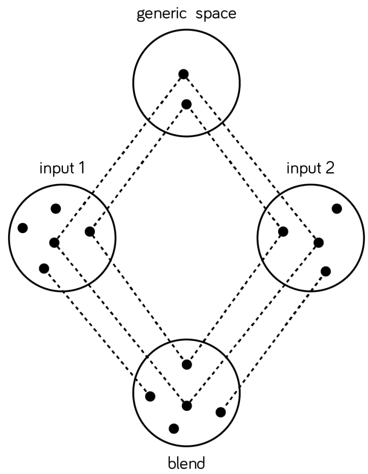


Figure 4.2: Classical model of Conceptual Blending.

used. Fauconnier and Turner (2002) identify four main types of integration networks, summarised below (Warchoł, 2018, p. 60):

- Simplex: only one input contains a frame, which is used to structure the blend;
- Mirror: Both inputs contain the same frame, which is used to structure the blend;
- Single-scope: Both inputs contain distinct frames, one of which is used to structure the blend;
- Double-scope: Both inputs contain distinct frames and the blend is structured by aspects of both input frames.

The *cross-space mappings* that occur between the two *input spaces* are based on what are referred to as *vital relations*, which include image schemas (e.g. PART-WHOLE), basic correlations (*cause-effect*), place, time, etc. (Fauconnier and Turner, 2002). Using these cross-space mappings, a fourth space is produced: the *blended space*. The blended space contains information projected from the two input spaces but also a new emergent conceptual structure, which is unpredictable from the input spaces and not originally present in them.

The projection of elements occurs in a selective manner, as not all elements from the input spaces are present in the blended space. Three processes are considered to take place in the construction of the blended space (Fauconnier and Turner, 2002, p. 47-48): *Composition*, *Completion* and *Elaboration*. Composition consists in bringing conceptual content from the input spaces into the blended space. Completion is understood as retrieving additional elements that may be necessary for the blend. Elaboration is an open-ended process, often referred to as "running" the blend. In this process, the *consistency* and *correctness* of the blend are tested, which may result in the blend being enriched with information considered pertinent or interesting. The open-ended nature of this process relies on the fact that it can vary in terms of enrichment.

As Ungerer and Schmid (2006, p. 265) highlight, the open-ended nature of conceptual blending raises questions in terms of the limits of the emergent structure. In this sense, Fauconnier and Turner (2002) propose *governing principles* (also referred to as *optimality principles*) that can be used to guide the process of conceptual blending and improve the quality of the produced blend:

- Topology: concerns the preservation of the conceptual structure of the input spaces in the blend;
- Pattern Completion: using existing integrated patterns to complete elements in the blend;
- Integration: highlights the importance of producing a blend that functions as a whole. Ungerer and Schmid (2006) refers to it as a sort of conceptual "gestalt";
- Maximisation and Intensification of Vital Relations: the vital relations should be maintained and promoted in the blend;
- Web: when being manipulated as a unit, the blend should maintain the network of connections to the input spaces;
- Unpacking: ensures that one can still reconstruct in initial inputs from the blend;
- Relevance: elements present in the blend should be relevant.

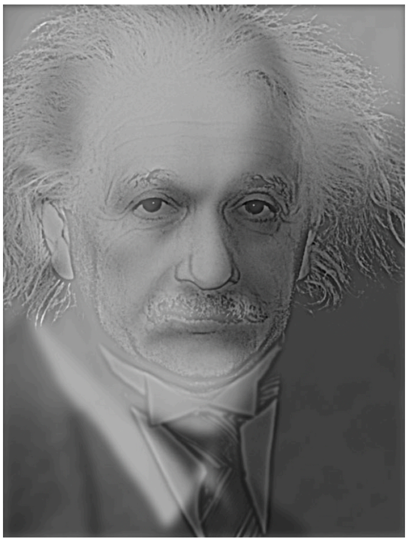
These principles are in competition with each other (Pereira, 2004) and their adjustment is crucial to make the blending happen, being responsible for the construction of meaning (Fauconnier and Turner, 2002). In this section, we only provide a brief description of Conceptual Blending. For more detail, we refer the reader to Evans and Green (2006), Fauconnier and Turner (2002), and Warchoł (2018).

The theory of Conceptual Blending has been used for different purposes, from the analysis of artworks (e.g. Warchoł, 2018) to the implementation of computational systems, e.g. *Divago* (Pereira, 2004). In this thesis, we explore how it can be used in combination with Visual Blending, proposing what we refer to as *Visual Conceptual Blending*.

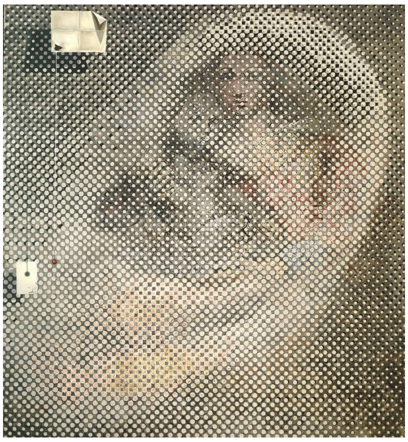
4.2 VISUAL BLENDING

Of the different ways of producing images, one can be referred to as *Visual Blending*. It can be defined as a process that consists in creating a given visual element from merging two or more existing ones.

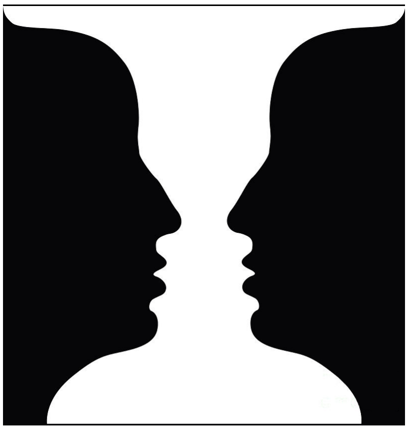
(a) *L'Amour de Pierrot*



(b) Hybrid image



(c) *Madonna*



(d) *Rubin's vase*

Figure 4.3: Double images: *L'Amour de Pierrot* by Salvador Dali, 1920 (top left); Hybrid image combining Albert Einstein and Sigmund Freud (Oliva, 2013) (top right); *Madonna* by Salvador Dali, 1958 (bottom left); and *Rubin's vase* (bottom right).

Figure 4.4: "Vertumnus" (ca. 1590) by Giuseppe Arcimboldo

From a broad perspective, *Visual Blending* can be considered to encompass images produced with different methods. One type is referred to as "double images" (Oliva, 2013). In these, distinct objects are perceived in a sequential manner, for example seeing different objects depending on the distance from the artwork or on whether the viewer is observing the image in a global way or focusing on local parts. One example is *L'Amour de Pierrot* (1920) by Salvador Dali (see Fig. 4.3a), in which elements in the composition are positioned in a way that a second image can be observed (e.g. a skull). Other techniques can be used to produce such images, for example by combining *low spatial frequencies* of one image with *high spatial frequencies* of another, resulting in what is referred to as "hybrid image" (Oliva, 2013) (Fig. 4.3b), in which the viewer is able to see a hidden image by squinting their eyes or moving away from the artwork. In a similar fashion, some images, referred to as "camouflage images", are produced with the goal of hiding one or more elements, which the viewer can disclose through attentive observation (Chu et al., 2010). A different technique consists in using dots of different colours and brightness to encode multiple contents, e.g. *Madonna* (1958) by Dali (Fig. 4.3c). Also worth mentioning are images referred to as "photomosaics", in which several images are used as tiles that construct a global image (Silvers, 1996; Xu et al., 2019). A similar approach is used with letters by Rebelo et al. (2018).

Another type of image that involves a *multistable perception* is referred to as "ambiguous images". In these, the viewer is presented with an image that has multiple interpretations, as is the case of *Rubin's vase* in Fig. 4.3d, which can be seen as a vase or two faces, or the *Duck-rabbit* in Fig. 4.5. Computational approaches have been developed for the production of such images, for example, Machado et al. (2015) employ an evolutionary strategy to produce ambiguous images.

In this thesis, we focus on a different kind of visual blending: images that result from the combination of existing ones but are perceived as a whole (i.e. a single object) and not to be observed in a sequential manner. This kind of visual blending is used in several fields from painting (e.g. "Vertumnus" by Giuseppe Arcimboldo, shown in Fig. **??**, and "Picnic in Central Park" by César Santos[1]) to animated films (e.g. the character Catbus in "My Neighbor Totoro" by Hayao Miyazaki). Although the simplest form of visual blending can consist in placing one image on top of another in an integrated manner, more complex processes involve the replacement of elements and even the use of image processing techniques, such as *Gaussian pyramids* (Burt and Adelson, 1983) and *Poisson* image editing (Pérez, Gangnet, and Blake, 2003).

Another area in which visual blending is often explored and studied is *Marketing*, being used to produce *visual metaphors*0. Despite this, most research is centred on the metaphor side and little attention is given to the visual blending process. In fact, visual blending is often ill-defined and can be confused with visual metaphor (e.g. Bolognesi, Heerik, and Berg, 2018; Peterson, 2018).

In this section, we provide an overview of visual blending. We start by clarifying the distinction between visual metaphor and visual blending. Then, we describe two main perspectives that take place when analysing visual blends: *structural* and *transformational*.

4.2.1 *Between Visual Blending and Visual Metaphor*

Despite being greatly explored, *Visual Blending* is often overlooked in existing research work, especially in studies of one of its potential uses: *Visual Metaphor*. For our study, it is important to make a clear distinction between the two terms.

Chilton, Petridis, and Agrawala (2019) state that a visual blend results from a process that merges two objects, each visually representing or symbolising an input concept. According to their definition, the two objects cannot simply be placed next to each other and should be integrated into a new one (the visual blend) in a way that they are still recognisable and allow the user to infer an association between the input concepts. This definition, as well as a great part of the research related to Visual Blending, comes from the context of *Advertising* and *Marketing*, specifically from studies that address how visual metaphor is used to communicate a given message more strongly. In fact, most studies focus on visual metaphor and the term "visual blending" is either poorly defined or completely absent. Even though our focus is on visual blending, we provide an overview of visual metaphor to make a clear distinction. It is important to mention that even though advertisements often use visual metaphor, a metaphorical interpretation is not always the goal and there are also many examples in which images are used as look-through[2] – see examples by McQuarrie (2008).

1 cesarsantos.com/santocesartnet/syncretism

2 Look-through pictures are meant to work as windows to what they depict; look-at pictures are not representations of objects but of ideas (McQuarrie, 2008).

Metaphor involves a cognitive process of viewing and understanding one domain (*target*) in terms of another domain (*source*), in which one or more features of the source are mapped on to the target domain, involving foregrounding, adoption or modification of certain features (Forceville, 2002a, p. 108; Peterson, 2018; Teng and Sun, 2002). In the context of metaphor, the term "domain" is used to refer to an entity or concept and the terms "source" and "target" to indicate the *directionality* (Peterson, 2018). A metaphor may occur in different modalities – e.g. *verbal*, *visual* or *gestural* (Ojha and Indurkhya, 2020). In the case of visual metaphors (also referred to as "pictorial metaphors"), the source and/or the target are visually represented in an image (Indurkhya and Ojha, 2017, p. 97). Several authors have studied visual metaphors in the past (e.g. Bolognesi, 2017; Carroll, 1994; Kennedy, 1982; Peterson, 2018; Zantides et al., 2016). Carroll (1994) suggests that visual metaphors are images composed of discrete elements that usually do not coexist (*noncompossible*), which are fused together to produce a *homospatially unified entity*. Carroll (1994) also highlights that the elements are to be perceived simultaneously – in contrast with images in which the elements can only be sequentially perceived, e.g. the duck-rabbit figure (see Fig. 4.5). Carroll's approach is analysed by Forceville (2002b), who refutes the requirement of *homospatially* and *noncompossibility* in a visual metaphor, mentioning that not all visual metaphors have a hybrid nature. Bolognesi (2017) acknowledges that different views on visual metaphor exist and formulates an encompassing definition in which "prototypical visual metaphors are (or better, can be found in) highly structured images that present perceptually-based incongruities that stimulate the viewers to construct cross-domain mappings to unravel the intended message". One key aspect of this definition is that it makes *incongruities*, defined by Schilperoord (2018, p. 19) as deviations from expectation, a central aspect of visual metaphor. Similarly, Schilperoord (2018, p. 42) states that anomalousness is a characteristic of images capable of inviting metaphor. However, metaphor is not a property of such images but instead the means by which the incongruity is resolved. In other words, one may say that there are no pictorial metaphors, only images that can stimulate a metaphorical state of mind Schilperoord (2018, p. 11). As such, the viewer attempts a figurative reading to make sense of the image (Cavazzana and Bolognesi, 2020) and visual metaphor occurs when the visual incongruity is resolved through a cognitive process based on a comparison between two terms. In any case, visual metaphor belongs to the denotative level and highly depends on cultural knowledge to occur – two people may interpret the same image differently.

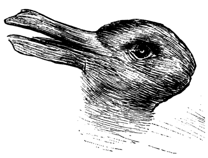


Figure 4.5: Duck-rabbit illusion

The reason for confusion between visual metaphor and visual blending resides in the idea that the two completely overlap, e.g. Petridis and Chilton (2019) state that Visual Metaphors "visually combine objects in an image in order to compare one to the other". Visual Blending –

defined as a process by which two (or more) visual objects are merged together, creating an image that may be perceived as incongruous – can produce images that may be interpreted as visual metaphors. Despite these being the sort of visual metaphor that is within the scope of our study, visual blending is not the only mechanism used to create incongruities – e.g. *one-domain incongruities* as described by Schilperoord (2018) do normally not involve a visual blending process. In the same way, not all visual blends are meant to be interpreted as a visual metaphor. For example, an image of a *centaur*, which can be seen as a visual blend between a *human* and a *horse*, is not necessarily meant to be interpreted as a visual metaphor but instead as a representation of a fictional organism – this is also pointed out by Cavazzana and Bolognesi (2020) who give the example of an *angel*. Other examples can be observed in the use of visual blending in icon production, in which the goal is a concrete interpretation, e.g. an iconic language for medical purposes by Lamy et al. (2008).

Another aspect that contributes to confusion between visual blending and visual metaphor resides in the partial overlap in terms of *structure*. Categorisation has been one of the topics addressed by researchers working with visual metaphor, leading to the proposal of several taxonomies to identify different types of visual metaphors.

4.2.2 *Visual Structure*

Forceville (2002a) proposes four subtypes of visual metaphor: (i) *pictorial similes*, in which both terms are fully present in a juxtaposed way; (ii) *MP1* or *contextual metaphors* (Van Mulken, Le Pair, and Forceville, 2010), in which one of the terms is absent but can still be inferred from the context; (iii) *MP2* or *hybrid metaphors* (Van Mulken, Le Pair, and Forceville, 2010), in which the two terms are fused into a single entity; and (iv) *verbo-pictorial metaphors*, which include elements from different modalities. The fourth subtype is not within the scope of our study, as we focus on monomodal examples (*visual mode*).

Phillips and McQuarrie (2004) propose a typology based on two *dimensions*: *visual structure* and *meaning operation*. The former is related to how the two domains are processed cognitively and involves two *types of operation*: *connection* and *comparison*. As our focus is on visual blending as transformation, meaning operation is the least important dimension of the two. On the other hand, visual structure is key when studying visual blending. Phillips and McQuarrie (2004) identify three types of structure, which are in line with the ones proposed by Forceville (2002a): *juxtaposition*, in which the source and target are depicted as complete entities; *fusion*, in which the two domains are combined to create a hybrid entity; and *replacement* in which one of the domains is omitted and only suggested through context. Phillips and McQuarrie (2004) frame their typology as an exhaustive list of ways of combin-

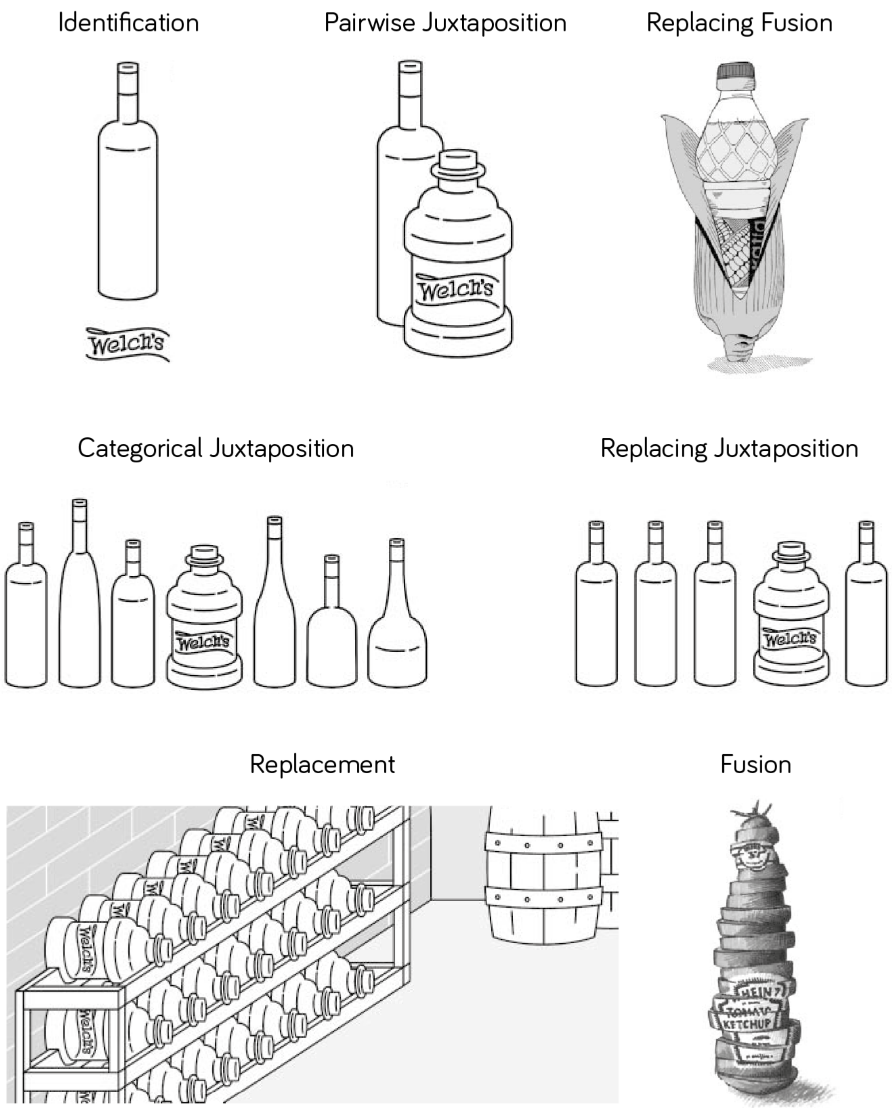


Figure 4.6: Taxonomy by Peterson (2018). Adapted from: (Peterson, 2018).

ing two image elements and state that no other possibilities need to be taken into account, as they would be either subcategories or a combination of the existing ones. Despite this, the typology is extended by McQuarrie (2008) by adding three additional types: *inclusion*, in which an element includes another element; *combination*, in which the two elements are combined to create a third; and *removal*, which is "one element without its expected complement". These new kinds of structures are not given any visual example and are only described textually. This aspect is also reported by Peterson (2018), who refers that the extension is not described in sufficient detail.

Peterson (2018) states that the main issue with the typology proposed by Phillips and McQuarrie (2004) is that it lacks intermediate types. To address this issue, Peterson (2018) makes a different proposal for extending the typology (Fig. 4.6), consisting of the following types:

- *identification* – one domain is represented pictorially and another domain textually;
- *pairwise juxtaposition* – presents two entities completely and separately, equal to juxtaposition by Phillips and McQuarrie (2004);
- *categorical juxtaposition* – the source entity is placed amidst a target set;
- *replacing juxtaposition* – one entity breaks a set of selfsame entities, replacing one instance;
- *replacement* – one entity is absent and must be imagined by the viewer using contextual cues, equal to replacement by Phillips and McQuarrie (2004);
- *replacing fusion* – part of one entity is replaced by another entity or part of it;
- *fusion* – two entities are fused to form a hybrid.

With this extension, Peterson (2018) aims at resolving issues from the typology by Phillips and McQuarrie (2004). For example, Peterson (2018) separates the idea of replacement (*replacing fusion*) from the absence of one domain (*replacement*), whereas Phillips and McQuarrie (2004) merge the two in *replacement*. Peterson (2018) admits that their typology may not settle all the existing debate and that future work may lead to improvements (e.g. some types may even collapse). Nonetheless, from the existing typologies based on structure, we consider the one by Peterson (2018) to be the most well-suited to our study of visual blending. Other proposals exist (e.g. Gkiouzepas and Hogg, 2011) but their analysis is not within the scope of this thesis and we refer the reader to Maes and Schilperoord (2008) and Peterson (2018) for further details on the topic.

4.2.3 *Perceptual features*

In addition to *structural analysis*, several authors have studied the importance of *perceptual features* when producing visual blends. For example, Chilton, Petridis, and Agrawala (2019) identify *shape similarity* as a common abstract structure in visual blends and develop a workflow for visual blending that produces blends based on user annotations of object shapes. As follow up, Chilton, Ozmen, and Ross (2020) present a study on how artists use identifying elements to create blends. Their results showed that the artists use three main visual dimensions to blend objects: *colour/texture*, *silhouette* and *details*. Based on these conclusions, they describe a visual blending system that takes as input two images that have been previously determined to have *shape similarity*. Then, it uses other features (*colour*, *texture* and *internal details*) to improve the blend by allowing the user to select them from the two initial objects.

One of the aspects that we can analyse from the work by Chilton, Ozmen, and Ross (2020) and Chilton, Petridis, and Agrawala (2019) is how shape is used to drive the process of visual blending – blends are produced by matching elements of similar shape. The importance of shape in establishing a connection between objects has been explored by multiple people. Schilperoord, Maes, and Ferdinandusse (2009) show how *associative relations* between two objects are often facilitated through Symmetric Object Alignment (SOA). SOA is defined as a design pattern that creates a *perceptual alignment* using two types of factors: *object-constitutive* (sensory attributes, e.g. *shape*, *size*, *colour* or *texture*) and *object-depictment* (how the object is displayed, e.g. *perspective* or *orientation*). The study by Schilperoord, Maes, and Ferdinandusse (2009) highlights how these factors can be used to increase the similarity between objects – e.g. an inverted guitar has a similar shape to a mushroom cloud caused by a nuclear bomb – which invites the viewer to connect the depicted objects. Ortiz (2010) contributes to this topic by analysing how SOA is often based on primary metaphors, e.g. SIMILARITY IS ALIGNMENT or THE NATURE OF AN ENTITY IS ITS SHAPE. Van Weelden et al. (2011) present a study on the impact of shape on a task of comparing objects placed side by side and their results further support the view that shape similarity helps the viewer in linking the objects.

Indurkhya and Ojha (2013) employ an algorithm that detects similarity between images based on *low-level features* (e.g. *colour*, *shapes*, *texture*, etc.) to study the effect of *perceptual similarity* on creative interpretation and visual metaphor understanding. Their results suggest that similarities attract attention and may work as anchors to the viewer's imagination in their search for *conceptual similarity*.

All these studies focus on object alignment in images placed side by side, which can be seen as a type of *juxtaposition*. Despite highlighting the role of perceptual features (among which shape) in stimulating a mental state of comparison, these studies do not fully align with the purpose for which shape is used by Chilton, Ozmen, and Ross (2020) and Chilton, Petridis, and Agrawala (2019) – the ground for exchanging elements when producing a hybrid object. This gap is filled by Indurkhya and Ojha (2017), who study *directionality* in visual metaphors and not only analyse *alignment* in juxtaposed images but also in cases of *replacement* and *fusion*. According to Indurkhya and Ojha (2017), in visual metaphors the images include hints or anchors that point to the intended message – these can be for example text or even the logo of the brand. Another type of hint that they refer to is the depicted shape or colour – one example given is how a curled up bottle of water suggests a tube of toothpaste (Indurkhya and Ojha, 2017, p. 109). In this case, shape is not only used for alignment but to allow the recall of the replaced element. This type of strategy is studied by Cavazzana and

Bolognesi (2020) who analyse examples of *contextual visual metaphor* (*replacement*) and describes mechanisms used to stimulate comparison. Cavazzana and Bolognesi (2020) refer to the concept of *occlusion shape*, which is defined by Hyman (2006, p. 76) as the shape of the mark one would need to make on a sheet of glass placed between the observer and an object to fully hide the object. Cavazzana and Bolognesi (2020) identify that *occlusion shape alignment* is used to hint that two things have similar features. In *replacement*, the replacing object is often modified (e.g. in size) to fit the shape of the replaced object. The sharing of the occlusion shape triggers the mental imagery of the viewer, causing the recall of the replaced element.

Another example given by Indurkhya and Ojha (2017) is a fusion between *hands* and *pliers*: the hands replace the tip part of the tool. Even though the hands do not have an exactly matching shape, they are positioned in a way that they recall the part that they replace. This last example is the sort of mechanism used by Chilton, Ozmen, and Ross (2020) and Chilton, Petridis, and Agrawala (2019) – choosing a replacement that is similar in shape.

As such, it is possible to identify different purposes of shape similarity depending on the type of blend: in *juxtaposition*, the alignment of shape in objects placed side-by-side stimulates comparison and increases the chances of metaphorical interpretations; in *replacement*, similar shapes are used to facilitate the recall of the replaced element, which helps the viewer to construct the message; and in *fusion* it not only facilitates the recall but also allows for a higher coherence among the object's parts, helping the viewer to perceive the hybrid object as a whole (a better merged image). Although we are focusing on *shape*, the same may occur with other perceptual features, e.g. *colour*. In fact, colour alignment is often used for producing coherent and realistic blends, especially when using *fusion* – see for example the analysis of visual blends by Martins et al. (2015) (also mentioned in Chapter 15).

In summary, perceptual features can be used at two different stages of the blending process. First, they can be used as a driver of the blending process, providing aid when deciding which elements should be linked and, depending on the type of blend, exchanged, based on their characteristics. Then, perceptual features can also be used to make elements seem similar by applying transformations, e.g. to make a replacement element better recall the replaced one.

4.2.4 *Transformational perspective*

As we have seen in the previous section, finding elements with shared properties makes it easier to not only establish a connection between them when using them on a *juxtaposition* blend but also to facilitate the exchange process in *replacement* and *fusion*. Nonetheless, it often occurs in visual blending that transformations occur beyond a simple exchange of parts, for example, to make objects seem similar, as we have previously noted.

One issue with structural classifications is that their categories are often too broad and include too many different types of transformation. Schilperoord, Maes, and Ferdinandusse (2009, p. 158) mention how some of the authors of existing classifications admit to the possibility of subcategories but consider them as irrelevant for not affecting consumer response – this is the case of Phillips and McQuarrie (2004) regarding horizontal and vertical arrangements in *juxtaposition*. Despite providing some insight into perception impact, a purely structural classification is too high level in the way that it does not exactly explain how the blending process occurs *transformational-wise*. Schilperoord, Maes, and Ferdinandusse (2009, p. 170) point out how in the case of *juxtaposition* a distinction is not made in terms of perceptual variation, encompassing several types of transformation, for example, SOA. This is an issue because, even within *juxtaposition*, different transformations (e.g. ways of positioning) may lead to different interpretations. The proposal by Peterson (2018) solves some of the issues related to *juxtaposition* from the categorisation by Phillips and McQuarrie (2004), for example by distinguishing juxtapositions of *one-vs-one* and *one-vs-many*. However, other issues have not been addressed, for instance, in the case of *one-vs-one* juxtaposing, the goal is often to stimulate comparison but it may also be to achieve a meaning based on *compositionality* – e.g. when juxtaposing a comics character with a modifier element. This lack of detail in regards to change mechanisms is especially problematic in the case of *fusion*, in which the blending process often resorts to several transformations.

On the other hand, the task of defining new categories is also a challenge. For example, by observing the three additional types proposed by McQuarrie (2008) – *inclusion*, *combination* and *removal* – one may argue that they are closer to a *transformational perspective* than to a *structural* one. We consider that the main difference between the two perspectives resides in their focus. While perspectives based on *visual structure* usually consider two domains as input and focus on how they are represented in the image, in *transformational approaches* the input are two visual representations and the focus is on identifying the necessary visual transformations to produce a merged image. The latter approach is the most useful when implementing a computational system for visual blending.

This can be observed in the recent approaches followed by authors working on visual blending: they are not so much focused on structure but rather on specific transformations. In the system presented by Chilton, Ozmen, and Ross (2020), after an initial shape matching procedure, transformations are used to improve the blend, such as *colour change*, *silhouette change* and *addition of details*. Similarly, Cavazzana and

Bolognesi (2020) mention that when taking advantage of *occlusion shape similarity*, elements are often changed in size to produce a better fit.

Ye et al. (2019) present a taxonomy of object transformations, which resulted from the analysis of an initial subset of 42 images and was then used to annotate 4,064 images. Their taxonomy is the following:

- *one type of texture* – a textured is borrowed from another object;
- *texture from separate objects* – a texture is created through the combination of small objects;
- *object inside another*;
- *object with missing part*;
- *hybrid object* – an object is created with parts of other objects;
- *bent object* – a distortion in shape occurs;
- *liquid deformed object* – a distortion in a liquid shape occurs;
- *context replacement* – one object is placed in the context of another.

These categories are not mutually exclusive, thus a single image may be assigned more than one category. One aspect that may be considered is that not all the transformations seem to be on the same level: some are focused on perceptual features and element parts, while others concern the object as a whole. The latter is aligned with structural taxonomies presented earlier: *hybrid object* and *context replacement* can be considered similar to the *fusion* and *replacement* from (Phillips and McQuarrie, 2004). In fact, Ye et al. (2019) report that these two categories are the most used ones and, given that the categories are not mutually exclusive, we believe that different topics are treated as equal, mixing two levels of analysis into one taxonomy. Regarding the remaining transformations, it is easy to identify that there are two main types: *perceptual* (e.g. *texture from separate objects*) and *structural* (e.g. *object inside another*).

It is important to highlight that a transformational perspective does not solve all the issues identified in structural approaches but may provide a more objective way of comparing different types of visual blends. We consider that the taxonomy by Ye et al. (2019) is a comprehensive list of transformations. In Chapter 5, we return to this topic and describe a study focused on visual blending transformations.

4.3 REQUIREMENTS FOR CONCEPT VISUAL REPRESENTATION

As addressed in Chapter 3, different computational approaches can be used to visually represent concepts. In this thesis, we explore the visual representation of concepts through blending. The idea of representation through combination is not novel, being a common method to produce vocabulary in visual languages. Horton (1994) makes reference to it in the following passage on how to produce icons that visually represent a given concept:

> *Draw a simple image already associated with the concept you wish to represent. If you cannot do this, then: represent the concept by combining familiar images in a simple way.*
> (Horton, 1994, p. 19)

In the previous sections, we have described how concepts can be combined to achieve different meanings and how images can be combined to produce novel representations. On the one hand, *Conceptual Blending* was introduced as a way to produce novel ideas. On the other hand, *Visual Blending* was introduced as a process in which several techniques can be used to combine images into a single output. The connection between the two levels – *visual* and *conceptual* – has been explored in the past, for example in the use of conceptual blending for the analysis of posters (Warchoł, 2018). We argue that the combination of the two levels can be explored for the purpose of visual representation of concepts. The work described in this thesis is focused on the exploration of this combination.

For the implementation of a system that visually represents concepts, we identify a set of necessities that need to be addressed. Some of them are related to the conceptual level, others to the visual and the remainder to how the two levels can be connected. In this section, we will describe what we consider key aspects for the implementation of a visual blending-based system for the visual representation of concepts.

4.3.1 *Conceptual Exploration*

As we have seen in Chapter 2 (Section 2.2), one of the strategies for conceptual reasoning is related to the extension to other concepts. For example, if one does not know what a *Xoloitzcuintle*, they can be enlightened by the description that a *Xoloitzcuintle* is a type of *dog* (i.e. a Mexican hairless dog). On the other hand, if one has to visually represent a general concept like *flower*, they will most likely resort to a specific flower, such as *rose*. In this sense, one of the requirements for a system that visually represents concepts is the ability to explore the conceptual space and make use of links between concepts.

This aspect is mentioned by Featherstone (2019, p. 38), who states that work on combinational creativity requires a knowledge base as input. Examples of knowledge bases are semantic networks of common sense knowledge, such as *ConceptNet* (Liu and Singh, 2004; Speer and Havasi, 2012). Moreover, the system is also required to be able to establish connections among concepts in a "useful" manner (Featherstone, 2019). This ability to establish and explore connections is the foundation for what we refer to as "conceptual extension" – representing a given concept by resorting to others.

4.3.2 *Visual Knowledge*

The capability to explore the conceptual space is important for the generation of visual representations but there is yet another aspect that is essential to consider. We may be able to extend a given concept to others to allow its representation. However, the limits of the system's representation potential depend on its visual knowledge. For example, a system may extend the concept *animal* to *dog* but it will still not be able to visually represent it if there is no visual representation or visual attributes related to *dog*. Thus, for the visual representation of concepts, the requirement of a knowledge base is not only related to the conceptual level but also to the visual one – a source of knowledge of visual properties is also a requirement for the production of visual artefacts.

4.3.2.1 *Conceptual Reach*

The sort of knowledge required obviously depends on multiple factors. First, it depends on what sort of content is to be represented. For example, the visual representation of emotions is explored by Lopes, Cunha, and Martins (2020) and of actions by Cruz, Hardman, and Cunha (2018). Second, it depends on what sort of visual representation one wants to achieve. As we have seen in the Chapter 2, there are many aspects involved in the production of representations of concepts – from the use of perceptual features, such as colour or shape, to the combination of more complex elements, such as icons. In any case, a main limitation of the system will be on the conceptual reach of the visual knowledge. For this reason, the system should be fed with a considerably large knowledge base to make it conceptually flexible (e.g. use a large dataset of input images).

4.3.2.2 *Structural Flexibility*

Another limitation related to the knowledge base on visual properties has to do with the type of data. Different types of data involve different processing methods. For example, if a system uses a dataset of colour meanings, its use in a visual blending system is low on complexity. On the other hand, if a dataset of raster images is used, image processing techniques will be needed for a visual blending process to take place. In our work, we take advantage of Scalable Vector Graphics (SVG) datasets, which facilitate the process of blending (more detail will be given in the following chapters).

4.3.3 *Visual-conceptual Alignment*

Having mentioned aspects related to the *conceptual level* and to the *visual level*, one last topic concerns the connection between the two. Ideally, in a system for the visual representation of concepts, the conceptual side and the visual side should influence each other. On the conceptual side, a system may search for ways to represent a given concept by resorting to others but, at the same time, the system should be aware of its visual knowledge limitations. On the visual side, the system should be able to use conceptual information to guide the process of representation, in our case the process of visual blending.

In a context of visual blending for the visual representation of concepts, an ideal alignment between the two levels would require the visual knowledge base to be previously complemented with semantic information (e.g. an element representing a *head* should be assigned with information that allows the system to know that it represents a *head*). This way, the system would be based on an integration of both levels, in line with what we refer to as *Visual Conceptual Blending*. Briefly, the basic notion of visual conceptual blending consists in conducting a visual blending process with a strong conceptual grounding, in which the conceptual level works as a guide and contributes to the overall quality of the final blend. We will return to the topic of Visual Conceptual Blending throughout the thesis, especially in Chapter 15.

4.4 SUMMARY

In this thesis, we focus on the visual representation of concepts using blending. Our approach can be divided into two different levels: *conceptual* and *visual*. In this chapter, we introduced these two levels.

We started by describing conceptual combination and presenting the different aspects of *Conceptual Blending*. Then, we shifted our attention to the visual side and provided a description of *Visual Blending*, mentioning what it encompasses and the different perspectives that should be considered when analysing or producing visual blends. In the last part of the chapter, we combined the two levels and identified the aspects that should be taken into consideration when implementing a blending-based system for the visual representation of concepts.

5 A CORPUS-BASED STUDY ON VISUAL BLENDING

In the previous chapter, we provided an overview of *Visual Blending*, *Visual Metaphor* and existing related research (Section 4.2). Although research on these topics is often framed within the field of *Marketing* and *Advertising*, our interest does not reside in these fields, focusing neither on ads nor on visual metaphor, but mainly on visual blending, specifically how visual blends can be produced.

As the main motivation behind our work is the implementation of computational systems for visual blending, we consider the collection of knowledge on the visual blending process to be a requirement. Such knowledge can be employed in the development of systems that produce visual blends or aid the user in their production. Our focus resides on studying how the process of visual blending occurs, specifically the set of techniques that are used.

Following the analysis of existing categorisations conducted in the previous chapter, we now propose a taxonomy of transformations that can be used in the production of visual blends. Then, we use the proposed taxonomy in the analysis of two image datasets that include visual blends (*VisMet* and *Emoji Kitchen*), focusing on the different types of transformation used. The two datasets are different in nature: *VisMet* is a dataset focused on visual metaphors, which includes both photorealistic and non-photorealistic images; and *Emoji Kitchen* is a non-photorealistic set of images produced through the combination of emoji with no intended meaning.

Our goal is to identify strategies that can be used when implementing computational systems for the generation of visual blends. Our study aims to explore two main questions:

- To what extent is it possible to identify the transformations often used in the production of visual blends and derive general transformational patterns?
- How can our findings be used in a computational system for the generation of visual blends?

With this chapter, we aim to contribute to the study of visual blending by focusing on the different transformations that can be used to produce blends.

5.1 CONTEXT

The process of visual blending encompasses several aspects that affect how a given blend is produced. On the one hand, visual blending can be based on a *perceptual similarity* between the objects to be merged – e.g. on *shape similarity* as identified by Chilton, Petridis, and Agrawala (2019). On the other hand, when conducting the blend a set of transformations are usually applied to the objects.

In Section 4.2.4, we described the existing approaches to the categorisation of visual blends. We argue that for the generation of visual blends, an approach based on transformations is best suited. Despite this, existing approaches are not without issues, some being focused on structure rather than transformations and others mixing different types of analysis (e.g. *perceptual* vs *structural*).

In addition, the study of transformations in visual blending is also limited, being addressed by few authors. One example is Ye et al. (2019), who conducted a study and identified seven different transformations used in visual blending. Despite showing the relevance of a transformational perspective in visual blending, we identify certain limitations in their study. First, the study by Ye et al. (2019) entirely focused on ads, covering 38 subjects (e.g. *food products*, *cosmetics*, etc.). Studying ads is advantageous when you are addressing visual metaphor, which assumes that there is a clear message and intention behind the blend. However, visual blending can also be used without an intended meaning. Although meaningless visual blends are not aligned with the topic of this thesis, from a transformational perspective their study is relevant to obtain a general understanding of visual blending operations. In addition, Ye et al. (2019) does not characterise the type of image that is studied but, from the examples provided in the paper, we assume that most of the ads are photorealistic. We are also interested in analysing blends that are produced in a non-photorealistic context.

5.2 TAXONOMY OF TRANSFORMATIONS

In Section 4.2.4, we provided an overview of existing taxonomies. Even though they are important contributions to the field, we consider that none of them is fully adequate to the study of visual blending from a transformational perspective.

Our work builds on existing work (Chilton et al., 2021; Peterson, 2018; Schilperoord, 2018; Ye et al., 2019). Table 5.1 shows the correspondence between taxonomies by other authors and ours.

Even though we base our work on existing taxonomies, the correspondence is not one-to-one in many cases, as we consider that some of the transformations are either too broad and abstract or too specific. This broadness is even acknowledged by some authors, e.g. Schilperoord (2018) describes *Distort* as an umbrella term that encompasses operations that produce visual incongruities by distorting an element's expected appearance. However, to be used for analysis purposes focused on transformations, the categories should be more concrete and limited. There are also two types that we considered too abstract in

Table 5.1: Comparison of Visual Blending Taxonomies. The table shows the list of categories and identifies similar categories from other authors' taxonomies – *Peterson* for Peterson (2018), *Schilperoord* for Schilperoord (2018), *Ye* for Ye et al. (2019) and *Chilton* for Chilton et al. (2021). The categories are divided into two groups: structural (S1-S6) and transformational (T1-T18). There are two cases marked as "x" as the work by Chilton et al. (2021) considers the transformation but does not name it. On the bottom are identified two categories that are not represented in our categorisation.

	Categories	**Peterson**	**Schilperoord**	**Ye**	**Chilton**
S1	Pairwise juxtaposition	Pairwise juxtaposition			
S2	Replacing juxtaposition	Replacing juxtaposition			
S3	Categorical juxtaposition	Categorical juxtaposition			
S4	Replacement	Replacement		Context replacement	
S5	Replacing fusion	Replacing fusion			
S6	Fusion	Fusion	Merge	Hybrid object	
T1	Add contents				Apply details
T2	Add element		Insert		Apply details
T3	Object inside object			Object inside Object	
T4	Object with missing part		Erase	Object with missing part	
T5	Replace part with object	Replacing fusion			
T6	Replace part with part	Replacing fusion	Substitute		
T7	Merge element				
T8	Colour replacement		Distort		Color
T9	Shape replacement		Distort	Bent object, Liquid deformed object	Silhouette
T10	Texture from separate objects		Distort	Texture from separate objects	
T11	Texture replacement		Distort	One type of texture	Texture
T12	Keep colour				x
T13	Keep contents				
T14	Keep shape				
T15	Keep texture				x
T16	Rotation		Distort		
T17	Scale		Distort		
T18	Flip				
	–		Transmute		
	–		Ambiguate		

terms of structure and transformation to be included in the taxonomy: *Transmute* and *Ambiguate* by Schilperoord (2018). *Ambiguate* is framed as an extreme case of *Merge* in which a given element can look like two different things (e.g. a white shape can be viewed as a sink or a tissue) and *Transmute* as the metamorphosis of one object into another. In the end, Schilperoord (2018) questions whether these need to be distinguished and for us it is not clear in terms of transformations how one can produce images that fit these categories. For this reason, we do not consider *Ambiguate* and *Transmute* in our categorisation.

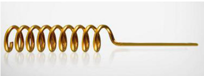


Figure 5.1: Straightened spring. Adapted: dataset by Ye et al. (2019).

Regarding the taxonomy by Ye et al. (2019), we also identified some issues. For example, we consider that the distinction between *Bent object* and *Liquid deformed object* is not necessary. We reached this conclusion by analysing examples from the dataset[1] provided by Ye et al. (2019). First, not all images in the dataset are visual blends – some of them portray an effect of a given product, e.g. a straightened out spring in an advertisement of an iron (Fig. 5.1). Then, most of the bent object examples are *shape replacement* (e.g. pizza in a radioactivity sign shape in Fig. 5.2), *texture replacement* (e.g. a floor with water waves in Fig. 5.3), or *texture with objects* (e.g. a spine made out of cups in Fig. 5.4). The same thing occurs in examples of *liquid deformed*: in most cases is a *shape* of a liquid being applied to an object. Taking the *transformational perspective* in which the inputs are the images to be merged, this is more a *shape replacement* than a *liquid bend*. For these reasons, we decided not to include these categories in our taxonomy.

Figure 5.2: Blend between pizza and radioactivity sign. Adapted: dataset by Ye et al. (2019).

Figure 5.3: Blend between water and street. Adapted: dataset by Ye et al. (2019).

Despite drawing inspiration from existing taxonomies, our proposal extends them by including new types of transformations, which were identified during our analysis. One type of transformation that we decided to include is related to the preservation of a property of a replaced element (i.e. property-preserving transformations), which is applied to the replacement element. This sort of transformation is used in the system presented by Chilton et al. (2021). In our taxonomy, we decided to divide it into four different transformations: *Keep colour*, *Keep contents*, *Keep shape* and *Keep texture*. Several other new transformations were included in the list.

Figure 5.4: Blend between cup and spine. Source: dataset by Ye et al. (2019).

Our proposed categorisation is divided into two levels of analysis: *structural* (s1-s6 in Table 5.1) and *transformational* (t1-t18 in Table 5.1). We consider that the proposal by Peterson (2018) is sufficiently detailed to allow its application. For this reason, the structural part of our proposal (s1-s6) is based on the taxonomy by Peterson (2018). However, an important distinction between our approach and the one by Peterson (2018) needs to be made. In the proposal by Peterson (2018), the *inputs* are considered to be *domains*, whereas we consider them to be *visual representations* (images). Peterson (2018) illustrates the different types of structures by using *grape juice* as the *target* domain and *wine* as the *source*, represented in most of the examples by a *Welch's* grape

1 people.cs.pitt.edu/~nhonarvar/data_analysis/interface.html

juice bottle and a wine bottle, respectively. Despite this, some examples show a glass of wine or a corkscrew instead of the wine bottle and are also identified as *pairwise juxtaposition*, mentioning that the source is still *wine*. This contrasts with a *transformational perspective* in which the glass of wine, the bottle and the corkscrew would be considered either different input objects or parts of a complex one containing all these elements.

In any case, by taking this distinction into consideration, one is still able to adapt the categories proposed by Peterson (2018) to a context of visual representations. The list of *structural categories* is the following:

- *Pairwise juxtaposition*: the two objects are placed together, e.g. *g* in Fig. 5.7;
- *Replacing juxtaposition*: replacement of a member of a set of elements, e.g. *a* in Fig. 5.8;
- *Categorical juxtaposition*: one object is placed next to a set of objects of the same category;
- *Replacement*: one object is placed in the context of another, e.g. *c* in Fig. 5.7;
- *Replacing fusion*: part of one object is replaced by the other object or part of it, e.g. *b*/*h* in Fig. 5.7;
- *Fusion*: hybrid object from the combination of the two objects, e.g. *a*/*d*/*e*/*f* in Fig. 5.7.

In regards to the *transformational* part of the taxonomy, it includes the transformations identified in our study. The list of transformational categories is the following:

- *Add contents*: elements contained by an object are added to the object, e.g. *c* in Fig. 5.9;
- *Add element*: element is added to other object, e.g. *a*/*e*/*g* in Fig. 5.7;
- *Object inside Object*: object is placed inside the other object, e.g. *g* in Fig. 5.7;
- *Object with missing part*: a part is removed from the object, e.g. *e* in Fig. 5.7;
- *Replace part with object*: one object replaces part of the other object, e.g. *c*/*h* in Fig. 5.7;
- *Replace part with part*: part of one object replaces part of the other object, e.g. *j* in Fig. 5.9;
- *Merge element*: two elements are merged together in a gradual transition, e.g. *f* in Fig. 5.7;

- *Colour replacement*: object is given the colour from another object, e.g. *d* in Fig. 5.9;
- *Shape replacement*: object is given the shape of another object, e.g. *f* in Fig. 5.9;
- *Texture from separate objects*: texture created from combining small objects, e.g. *d* in Fig. 5.7;
- *Texture replacement*: object is given the texture from another object, e.g. *b* in Fig. 5.7;
- *Keep colour*: object keeps initial colour;
- *Keep contents*: object keeps initial contents, e.g. *h* in Fig. 5.7;
- *Keep shape*: object keeps initial shape;
- *Keep texture*: object keeps initial texture, e.g. *h* in Fig. 5.7;
- *Rotation*: the object is rotated, e.g. *h* in Fig. 5.9;
- *Scale*: the size of the object is changed, e.g. *g* in Fig. 5.9;
- *Flip*: object is reflected, e.g. *a* in Fig. 5.9.

By observing the transformations identified, one may realise that different kinds exist, for example *structure-related transformations* (*Add element* or *Replace part with object*), *property-related transformations* (*Colour replacement* or *Texture replacement*) and transformations concerning the *object itself* (*Rotation* or *Scale*). It is important to mention that we do not consider the proposed list as a closed one, as we believe that other types may arise through further analysis. Moreover, a given blend may be produced by using several combinations of transformations. In our analysis, we choose the less complex combination.

5.3 ANALYSIS METHODOLOGY

Visual Blending is a process that can be used to produce different types of images. Our approach was to study and compare different datasets of images in which visual blending occurs. For this reason, we decided to analyse images from two different sources: *VisMet*, a dataset of visual metaphors; and *Emoji Kitchen*, a dataset of emoji blends.

We assume that in most cases the image may have a message behind it. Despite this, an analysis focused on meaning is out of scope for our study. We consider meaning construction as a step that starts before the transformational process that is the focus of our study. Moreover, our analysis would be based on interpretation, which is often complex, depending on the viewer and their cultural background. For an example of an analysis based on interpretation, we refer the reader to Poppi, Bolognesi, and Ojha (2020).

Instead, we centre our analysis on the transformations used to produce the visual blend from two initial input objects. In the scope of our work, we refer to the entire input as *object* and to the parts of the initial input as *elements* or *parts*. In a visual blend, it is often the case that one of the input objects is used as the *base* – the one that is most depicted in the final blend, despite being subject to replacements and additions – and the other is used as the *modifier* – usually having its parts, properties and even itself added to the final blend or used as a replacement. For example, in a blend between a face and a crown (see Fig. 5.5), the crown can be used as the modifier by being added to the top of the face's head shape, which works as the base.

Figure 5.5: Blend between *grinning face* `U+1F600` and *crown* `U+1F451`. Source: *Emoji Kitchen*.

Our methodology of analysis is based on a set of steps:

- identify images in which at least two different objects are merged and filter out images that are out of scope to our study (this is further explained in each analysis section);
- identify the type of structure. In our analysis, we will use letters to refer to each of the types: Pairwise juxtaposition (PJ), Replacing juxtaposition (RJ), Replacement (R), Replacing fusion (RF), Fusion (F) and Categorical juxtaposition (CJ);
- identify the transformations that were performed to produce the visual blends;
- identify if there is alignment between the inputs of the blend, which can be *conceptual* or *perceptual* (*shape*, *colour*, *texture*, etc.).

Even though we apply this methodology in the analysis of the two datasets, the nature of the *Emoji Kitchen* dataset allows for a more extensive study. First, the number of analysed blends is much higher in *Emoji Kitchen* than in *VisMet*. Second, in the case of *Emoji Kitchen* we have access to images that can be considered the initial input of the visual blending process, whereas in *VisMet* we only have access to the final blends, only allowing for a speculative analysis based on what we consider the initial inputs were. We describe both studies in the following sections.

5.3.1 *Dataset #1: VisMet*

VisMet is an online corpus of annotated images that are considered to show visual metaphors (Bolognesi, Heerik, and Berg, 2018). It was released in 2014 to be used by researchers interested in visual rhetoric. Considering that visual metaphors can be produced using visual blending, we decided to use the corpus as one of the datasets of our study.

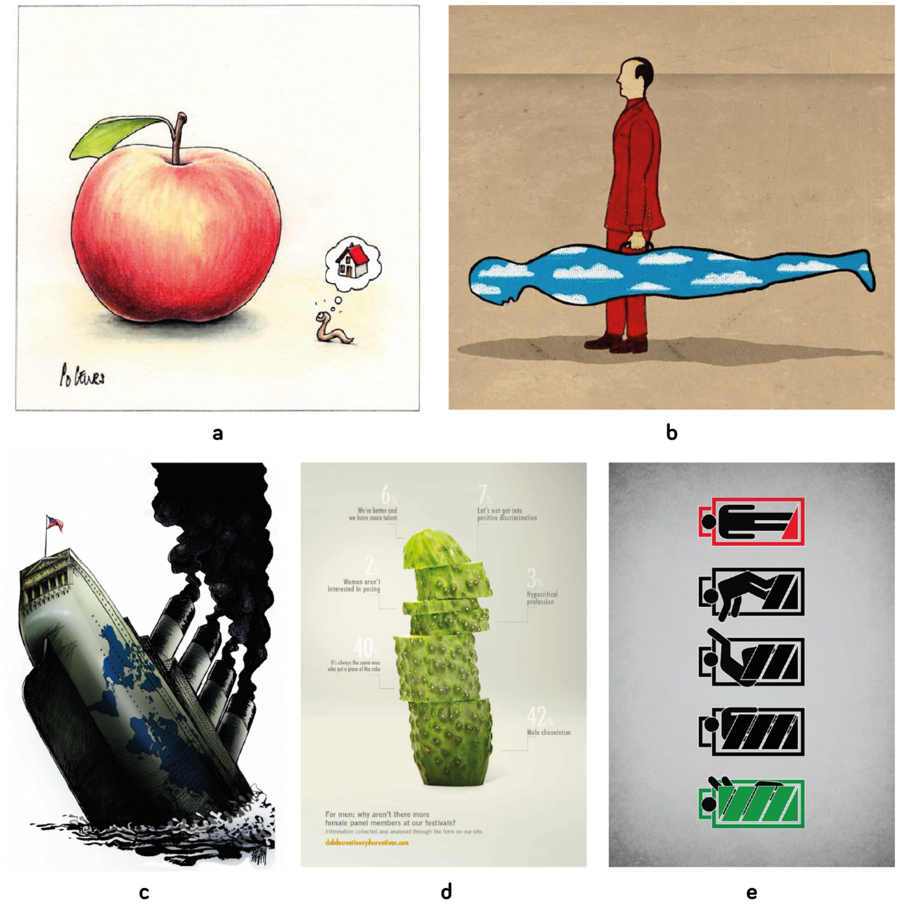


Figure 5.6: *VisMet* exclusion examples: *a* (by Pol Leurs) excluded due to *narrative* and *no blend*; *b* (by Beppe Giacobbe) excluded due to *blend sequence*; *c* (by Angel Boligan) excluded due to *composition* and *multiple blend*; *d* (by Saatchi & Saatchi) excluded due to *visualisation*; and *e* (by Viktor Hertz) excluded due to *composition* and *sequence*.

5.3.1.1 *Initial Analysis*

The corpus of *VisMet* consists of 353 annotated images (retrieved in March 2021), 299 of which were identified as non-photorealistic and 51 as photorealistic. However, not all images are suitable for our analysis. Firstly, as described by Bolognesi, Heerik, and Berg (2018), the corpus is composed of both images that present visual incongruences and images in which the incongruence is not fully perceptually but rather derived from a relation between linguistic elements (e.g. title) and the visual representation. Images belonging to the latter type are not within the scope of our study, which is focused on visual blending. Secondly, not all images with visual incongruences, considered visual metaphors, use visual blending. Thirdly, even though a given image makes use of visual blending, it may only be a small part of it.

As we intend to focus on images that use visual blending as the main production process and result in a composition with one main object, we conducted an initial filtering of the corpus. The reasons for exclusion are as follows:

- *composition*: the image is composed of multiple individual elements (e.g. *symbols*, *text*, etc.) that contribute to an overall single meaning (e.g. *c* and *e* in Fig. 5.6);
- *narrative*: the image follows a narrative approach, which is only obvious when analysing multiple objects (e.g. *a* in Fig. 5.6). These objects play roles in the narrative and the dimension of time is meant to be considered when analysing the image (there is a before and an after);
- *sequence*: the image follows a sequence structure, in which the depicted elements should be interpreted as a sequence of events (e.g. *e* in Fig. 5.6);
- *visualisation*: the image is a data visualisation (e.g. *d* in Fig. 5.6);
- *multiple blend*: the image shows multiple cases of the visual blending, i.e. more than two entities (e.g. *c* in Fig. 5.6);
- *blend sequence*: the image shows the result of a sequence of visual blending processes (e.g. a visual blend between two entities occurs, followed by a visual blend with a third one). Not to be confused with a visual blend in which different parts belonging to the two entities are blended (e.g. *b* in Fig. 5.6 is a sequence of briefcase → person → sky);
- *no blend*: no visual blending occurs (e.g. *a* in Fig. 5.6).

It is important to mention that a single image can be assigned more than one reason for exclusion. In total, 228 images were excluded – the majority of exclusions were due to the image being considered as a *narrative* (105) or a *composition* (87). As such, we analysed a total of 125 images: 29 photorealistic and 96 non-photorealistic.

5.3.1.2 *Structural Analysis*

We started by conducting a structural analysis of the visual blends. The most used types of structure are RF (46.4%), F (29.6%) and R (16.8%). PJ was only identified in 4.8% and RJ in 2.4%. When comparing the type of structure in photorealistic and non-photorealistic blends, most types vary only by 1-2% and the highest differences are found in F (37.9% in photorealistic and 27% in non-photorealistic) and R (10% vs 18.7%).

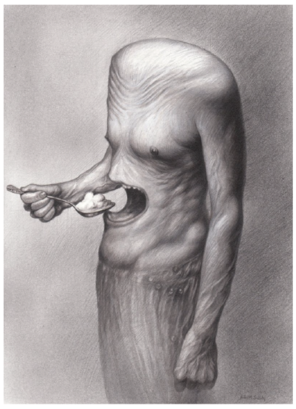
Another observed aspect is that some blends show an element or property being used multiple times, these occur in a blend when an element/object is used as replacement more than once or when a property (e.g. *colour*) is used in more than one element. We identified 33 blends with multiple uses, most of which were RF (19) and F (8).

a) Fusion
Add Element

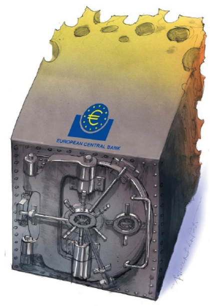
b) Replacing Fusion
Texture replacement

c) Replacement
Replace part with object

d) Fusion
Texture from separate objects



e) Fusion
Add element,
Object with missing part

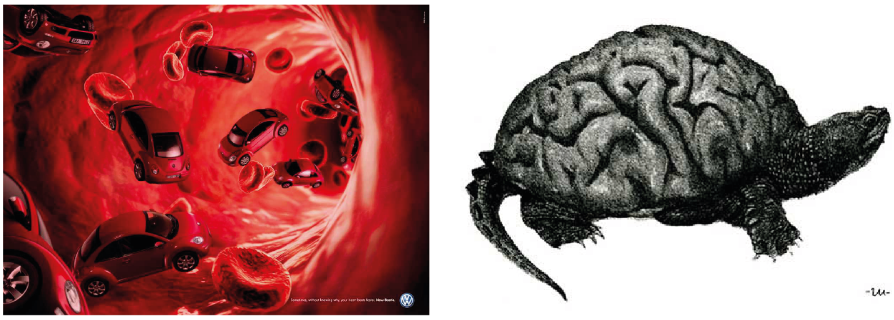


f) Fusion
Merge element

g) Pairwise juxtaposition
Add element,
Object inside object

h) Replacing Fusion
Replace part with object, Keep texture, Keep contents

Figure 5.7: *VisMet* examples, identifying the structural category and transformations, e.g. in (*a*) *Fusion* and *Add Element*, respectively. (**a**) a barred window is added to the sole of a boot (by Azizy Avuzdogan); (**b**) the texture of crackers is applied to the sofa (by Kolle Rebbe); (**c**) men replace nails (by Marcin Bondarowicz); (**d**) texture made of coffee beans is applied to waves (by Not Perfect, Y&R); (**e**) a mouth is added to the belly and the head is removed (by Agim Sulaj); (**f**) cheese merges with a vault (by Rainer Ehrt); (**g**) a paraglider is added inside a jar (by Paolo Dalponte); and (**h**) a building structure is replaced with a brief case but its texture and elements (windows and door) are kept (by Agim Sulaj). Adapted from: *VisMet*.



a) Replacing juxtaposition
Replace part with object

b) Replacing fusion
Replace part with object

Figure 5.8: Colour and shape alignment. Source: *VisMet*, by DDB Spain and Halász Géza.

5.3.1.3 *Transformational Analysis*

The main focus of our study resides in analysing the type of transformations used to produce the blends. As such, for each visual blend, we identified the set of transformations conducted in the process of blending. It is important to mention that in some cases the same blend can be produced using different sets of transformations. In our analysis, we identified the set that requires fewer transformations.

Given that no initial input images are available for *VisMet* and, for that reason, our analysis focuses on the final blend, some transformations are not possible to be identified. This was the case of *Flip* and *Scale*, which require a comparison with an initial state. *Rotation* is also difficult to assess but, in some cases, it is possible to identify (when the element is fully inverted in relation to its usual state). In any case, we consider that both *Scale* and *Rotation* are often used to allow a better placement of the elements.

The most used transformation is *Replace part with object*, being used in 50 blends (40% of the analysed blends) – 29 considered RF, 18 R and only three RJ. Moreover, it is the only transformation used in 36 blends. The second most frequent is *Add element*, used in 28 blends (20 of which are F). In third place are *Shape replacement* and *Texture replacement*, both used in 14 blends.

Three transformations that are used less frequently but still reach values around 7-8% of the blends are *Replace part with part*, *Merge element* and *Texture from separate objects*.

5.3.1.4 *Alignment Analysis*

Despite not having access to the original input, we annotated the images in which we could identify that there was alignment between the exchanged elements. In total, we were able to identify alignment in 59 blends. Most of the cases are either RF (32 blends) or F (21 blends). Four types of alignment were identified: *colour* in five blends (e.g. red car and red blood cell in Fig. 5.8); *conceptual* in 16 (e.g. matching a shopping car with a tank); *shape* in 46 (e.g. matching a flute with a screw or a turtle shell with a brain in Fig. 5.8); and *texture* in two (e.g. texture curtain with a sculpted column).

By only having access to the final blend, it is not possible to address topics related to the input objects. In any case, we believe that these results already shed some light on how transformations are used.

5.3.2 *Dataset #2: Emoji Kitchen*

In this section, we describe the analysis of the second dataset: *Emoji Kitchen*. *Emoji Kitchen* is a feature available in *Google*'s keyboard *Gboard* that allows users to merge emoji, presenting them with blends in the form of custom-made stickers. Even though these images are only available in messaging apps on *Android* devices, we were able to collect a subset (2243 images) by using the website *Emoji Menu*[2] – an unofficial searchable catalogue of *Emoji Kitchen* combinations developed by Ben Grant and Ben Koder.

Despite little information being available regarding *Emoji Kitchen*, from the description on its website one can conclude that the mashup design follows the same style as the open-source *Noto Emoji* font. By comparing images from Noto with the blends, we concluded that a total of 179 different emoji are used in the subset, from which 22 have subtle style changes but maintain the same overall look (as emoji are standardised, they remain very similar), ten used more than one version in the blend set (e.g. the monkey emoji has a version with only its head and one in full body) and only four can be considered sufficiently different (e.g. different configuration of the avocado emoji). As such, it is possible to compare the initial emoji with the produced mashup to assess the conducted transformations in most cases. This is advantageous as some transformations are only possible to be identified by comparing with the initial input (e.g. *Flip*).

5.3.2.1 *Initial Analysis*

Our goal is to conduct an analysis on visual blending based on input images and transformations used to produce a blend. As the *Emoji Kitchen* blends are produced by a designer, it is normal that there is freedom to elaborate beyond the initial input. An example of *elaboration* is a blend that includes arms, even though these are not present in the input emoji. Another example is the blend between the *face with head-bandage* 🤕(`U+1F915`) and *hot beverage* ☕ (`U+2615`). The blend shows a broken cup (see Fig. 5.10), which is an elaboration based on the idea of "injury" applied to a cup. Despite considering these as valid blends,

2 `emojikitchen.rocks`

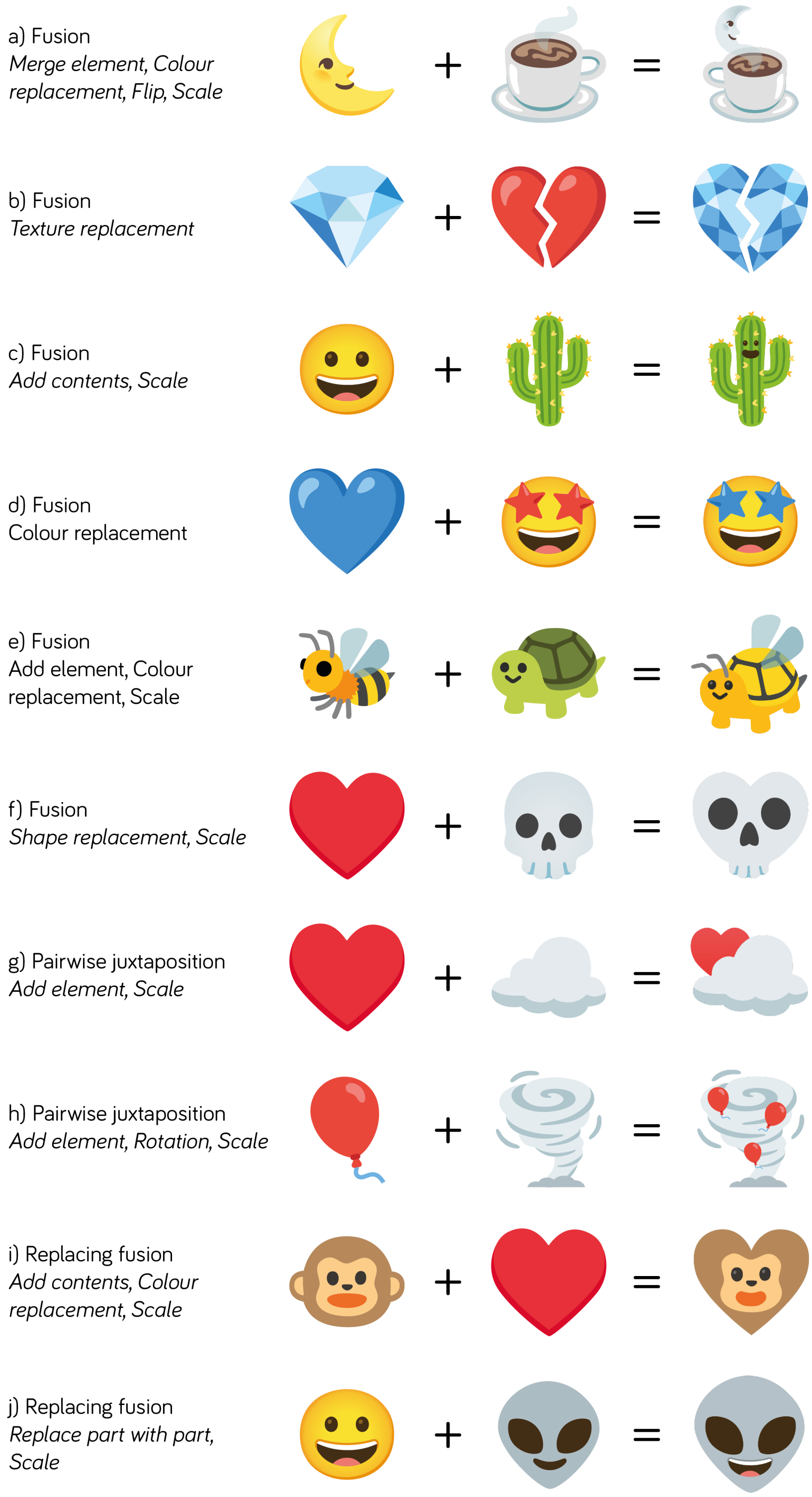


Figure 5.9: *Emoji Kitchen* examples, identifying the structural category and transformations, e.g. in (*b*) *Fusion* and *Texture Replacement*, respectively. Adapted from: *Emoji Kitchen* and *Noto Emoji*.

Figure 5.10: Blend between *face with head-bandage* (`U+1F915`) and *hot beverage* (`U+2615`). This blend was excluded due to elaboration. Adapted from: *Emoji Kitchen* and *Noto Emoji*.

they are out of scope for our study as they do not simply use transformations and resort to open elaboration.

Therefore, we excluded images from our *Emoji Kitchen* study due to two reasons:

- *elaboration*: the blend contains elements that are not present in any of the two input emoji (see Fig. 5.10). This sort of blending would require more information than just the initial input and is out of the scope of our study;
- *similarity*: the blend seems to be identical to an emoji.

A total of 137 blends were excluded due to *elaboration* and 18 due to *similarity*. This type of exclusion process is only possible as we have the *Noto Emoji* as a ground for comparison. After removing the excluded images, the final set had a total of 2088 images.

5.3.2.2 *Structural Analysis*

We started by conducting a structural analysis. Considering the 2088 not excluded blends, we could only identify three types of structure: PJ (20.45%), RF (55.75%) and F (23.8%). An interesting result is the absence of replacement (R), which is justified by the lack of context when combining emoji.

Taking advantage of having the *Noto Emoji*, which we consider as base input for the blends, we were able to annotate the basis for the blending process: *full object* (the blend was based on an exchange of a full object), *part* (the blend was based on an exchange of part of an object) or *property* (e.g. colour transformation). From our analysis all *property-based blends* were considered F, which consisted of 374 out of the 497 F blends; PJ mostly used full object exchanges (392/427); and in RF the majority of blends (657/1164) were part-based.

During our analysis, we also identified cases of multiple uses of a given element (e.g. balloon added three times in *h* in Fig. 5.9) or property (e.g. blue colour applied twice in *d* in Fig. 5.9). The majority of these cases were observed in PJ blends (87 out of 166 cases). Interestingly, 103 out of the 166 cases were blends based on an exchange of the full object, 40 based on property and only 15 on part. We also noticed six cases in which there was a one-for-many exchange in the blend (e.g. two eyes being replaced by a bigger eye, Fig. 5.11). Regarding RJ,

Figure 5.11: Blending *eye* `U+1F441` and *turtle* `U+1F422`. Replacing fusion using *Replace part with object*, *Scale*. Adapted from: *Emoji Kitchen* and *Noto Emoji*.

Peterson (2018) describes it as one object replacing a member of a set of identical objects. Despite not being able to find examples that fully align with this definition, we found five blends in which an object is used to replace multiple instances of identical parts of the other object (see Fig. 5.12) – these were labelled as RF. We also found cases in which the two objects were included fully in the blend but repetition would occur – either one or both of them are placed more than once (see Fig. 5.13). In any case, we do not consider these blends as RJ because the objects are not perceived as belonging to the same group but only work as a composition – we labelled them as PJ.

Figure 5.12: Blend between *gem stone* (`U+1F48E`) and *cupcake* (`U+1F9C1`). Source: *Emoji Kitchen*

Figure 5.13: Blend between *kiss mark* (`U+1F48B`) and *turtle* (`U+1F422`). Source: *Emoji Kitchen*

5.3.2.3 *Transformational Analysis*

After conducting the structural analysis, we focused on the transformations used to produce the blends. The most frequent transformation is *Scale*, which is used in 1607 blends. Despite this high number, we consider it not very relevant transformational-wise as changes in size are normally used to make the element better adapt to the new position (e.g. by fitting the space of a replaced part). Similarly, *Rotation* (309 blends) is usually employed for alignment purposes (e.g. using the same rotation of the replaced element).

The second most used transformation is *Colour replacement* (a total of 617) and no other transformation is used in more than half of the blends that resort to it. In terms of structure, 407 are labelled as F (it is the most frequent transformation in *fusion*), 199 as RF and only 11 as PJ.

Add element is the third most used transformation (a total of 553 blends) – used 360 times in PJ (188 of which used only with *Scale*), being its second most used transformation, 118 times in RF and 75 in F.

Replace part with object is the fourth most used transformation and 498 out of its total of 510 occurrences are labelled as RF, 354 from which used only with *Scale*. The transformation *Add contents* is also highly used in RF and in more than half of the cases it is used only with *Scale* (230 out of a total of 434). *Replace part with part* is also worth mentioning, as it is identified in 220 blends, 204 of which are labelled as RF. The remaining transformations are used less often and can be analysed in Table 5.4 in the discussion section.

5.3.2.4 *Alignment Analysis*

Another aspect that we considered when analysing the blends was the possible *alignment* between the exchanged parts, which could be seen as the basis for the mapping that guided the blend. We identified the different types of similarity that are used to produce the mappings:

- *conceptual*: the parts share a conceptual connection (e.g. head maps with head);
- *colour*: the parts have a similar colour;
- *shape*: the parts have a similar shape (e.g. a circular shape);
- *proportion*: the proportion of the parts is similar.

We were able to identify *alignment* in 1137 out of the 2088 blends (54.4%), 1085 of which were RF and 52 were F. Most of the cases were considered an alignment of proportion (588) – which makes sense as the blending process is less complex if the exchanged elements have a similar proportion, e.g. avoiding the necessity for high distortion to fill the replaced space. Having a similar shape is even better but it only happens in 312 blends. In third place is the conceptual alignment (233). Colour alignment, despite being seen as a way to produce good blends (Martins et al., 2015), only occurs in three blends.

5.3.2.5 *Individual Analysis*

One difference between analysing the *VisMet* dataset and the *Emoji Kitchen* has to do with the input visual representations. While in *VisMet* we do not have access to input objects and each blend is completely different, in the case of *Emoji Kitchen* the input elements (emoji) are available and all of them are used in more than one blend. This allows an analysis to be made in terms of input objects, which is not possible with the *VisMet* dataset.

The subset of *Emoji Kitchen* blends that were analysed is based on a total of 179 different emoji. Despite this, the emoji are not used evenly, reaching a maximum of 171 uses (*gem stone* 💎 `U+1F48E`), and a minimum of six (*smiling face with open mouth* 😃`U+1F603`) (standard deviation of 36.3, average of 23.3 and mode of 11).

Our main goal is to understand which transformations are used. First, as we already mentioned, *Scale* and *Rotation* are mostly used to improve the exchanges, having little relevance. As such, in the following analysis, we will pay little attention to them as transformations, even though they may be predominant for a given emoji. During our analysis, we noticed that the blends of certain emoji often follow similar approaches. For example, from the 160 blends that used the *newspaper* 📰 `(1f4f0)`, 41 blends used a strategy based on a combination of *Replace part with object* and *Colour replacement* – the other emoji is used

Figure 5.14: *Emoji Kitchen* individual strategies from *crown* 👑 `U+1F451` (*Pairwise juxtaposition* using *Add element, Rotation, Scale*), *crystal ball* 🔮 `U+1F52E` (*Pairwise juxtaposition* using *Object inside Object, Scale*), *hole* 🕳 `U+1F573` (*Pairwise juxtaposition* using *Add element, Scale*) and *newspaper* 📰 `U+1F4F0` (*Replacing fusion* using *Replace part with object, Colour replacement, Scale*).

as replacement and converted into grayscale (see Fig. 5.14). Given that the blends are produced by a designer, it is not unusual that such strategies are used. We believe that the analysis of these strategies can lead to interesting conclusions regarding the process of blending. As such, the first step was to identify frequent transformation combinations. To do this, for each emoji we identify combinations that are frequently used in its blends, using four groups based on percentage of blends that use the combination (*comb_use*):

- ">90%" ($comb\_use > 90\%$, i.e. combinations that are used in >90% of the blends of a given emoji);
- ">75%" ($75\% < comb\_use \leq 90\%$);
- ">50%" ($50\% < comb\_use \leq 75\%$);
- ">30%"($30\% < comb\_use \leq 50\%$).

The most identified combinations are: "Replace part with object, Scale" (65 emoji[3]), "Colour replacement" (43 emoji) and "Add element, Scale" (9 emoji). There are only two cases of combinations in the group ">90%" and three in the ">75%" (all corresponding to either "Colour replacement" or "Add element, Scale") but four of these cases are emoji that have less than ten blends each and the other has less than 20. The group ">50%" has a total of 23 cases and ">30%" has 119.

3 65 emoji have the combination in one of the identified frequency groups

Another type of analysis focused on how the emoji are used. For this, we tried to identify the *base* and the *modifier* for each blend. Although the *base* is often the emoji that is more represented in the blend, there are cases in which it is not easy to identify these roles. We were able to identify the *base* and *modifier* in 2024 blends.

The first aspect that we concluded is that there are certain emoji that are more prone to being *base* and others to being *modifier* – 13 emoji are used as a modifier in more than 75% of their uses and 24 emoji are used as a base in more than 75% of their uses (see Table 5.2). Two examples of typical modifiers are the *crown* 👑 (used in 168 blends, 95.8% as the modifier) and *gem stone* 💎 (used 170 times, 90.1% as the modifier). Two examples of typical bases are the *crystal ball* 🔮 (used 165 times, 97% as the base) and the *turtle* 🐢 (used 162 times, 85.3% as a base). However, information about whether an emoji is often a base or a modifier is not very useful on its own as, in addition to being emoji-specific, it does not provide any clues on which transformations are used – two emoji, one often base and one often modifier, may be combined in many different ways.

If we go back to the example of the *newspaper* emoji 📰 (`U+1F4F0`), in which 41 out of its 160 blends used a strategy based on the combination "Replace part with object, Colour replacement, Scale", one interesting result is that the newspaper emoji is considered the base in all of the 41 blends. To identify more of these strategies, we decided to conduct an analysis based on typical uses of the emoji (base/modifier) and on the number of blends that use the emoji. From the emoji that are base or modifier in more than 75% of their blends, we selected the ones that are used in more than 140 blends, which led to four modifiers and five base emoji. By focusing on individual emoji and transformations, there are two main ways of analysing the results: a general perspective of each transformation (e.g. blends that use *Colour replacement*) or by focusing on specific combinations of transformations (e.g. blends that only use *Replace part with object* and *Colour replacement*). While the former allows us to identify predominant transformations, the latter makes it possible to identify specific strategies.

As previously described, we identified frequent combinations for each emoji based on the percentage of blends in which they are used. From our selection of emoji, there are two emoji with a combination in the ">50%" group:

- *crown* 👑 (168 blends, 95.8% as modifier) uses the combination "Add element, Rotation, Scale" in 119 blends. If one considers all blends that use "Add element", the number rises to 154 blends. See examples of this combination in Fig. 5.14;

- *fork and knife with plate* 🍽️ (144 blends, 79.2% as base) uses the combination "Add contents, Scale" in 73 blends. If we consider

Table 5.2: Base and Modifier Emoji. The table shows the emoji that are used as base or modifier in more than 75% of the blends that have them as input. For each emoji, we show its *Noto* rendering, the code point, the percentage of blends that it is used as a base (%B) or modifier (%M) and the number of analysed blends that have it as input (#).

	Base				*Modifier*		
	code	%B	#		**code**	%M	#
🌟	U+1F31F	100.0%	10	💫	U+1F4AB	100.0%	9
🎗	U+1F397	100.0%	9	👑	U+1F451	95.8%	168
❣	U+2763	100.0%	19	💎	U+1F48E	90.1%	170
🔮	U+1F52E	97.0%	165	♥	U+2665	81.6%	157
🎊	U+1F38A	95.0%	20	💛	U+1F49B	80.0%	50
💖	U+1F496	94.7%	19	💚	U+1F49A	77.4%	52
🕳	U+1F573	94.7%	19	🤎	U+1F90E	77.4%	52
💓	U+1F493	90.0%	20	💜	U+1F49C	76.9%	51
💟	U+1F49F	89.5%	19	🧡	U+1F9E1	76.9%	51
🌹	U+1F339	88.9%	9	💙	U+1F499	76.5%	50
🎂	U+1F382	85.7%	7	🖤	U+1F5A4	76.5%	50
🐢	U+1F422	85.3%	162	😀	U+1F600	76.5%	152
💕	U+1F495	85.0%	20	🤍	U+1F90D	76.5%	50
☕	U+2615	84.8%	151				
💐	U+1F490	83.3%	18				
🎈	U+1F388	81.2%	164				
🌛	U+1F31B	80.0%	20				
🌜	U+1F31C	80.0%	20				
🌪	U+1F32A	80.0%	10				
🌷	U+1F337	80.0%	15				
💗	U+1F497	80.0%	14				
🍽	U+1F37D	79.2%	144				
💌	U+1F48C	78.6%	14				
🐙	U+1F419	76.9%	12				

all blends that use “Add contents”, the number rises to 93. We also see that many blends use “Add element” (44).

The other emoji from our selection have combinations in the “>30%” group. We describe them below:

- *crystal ball* 🔮 (165 blends, 97% as base) has two combinations in the “> 30%” group: “Add contents, Scale” is used in 71 blends (96 blends use “Add contents”) and “Object inside Object, Scale” in 62 (64 use “Object inside Object”, which is more than 95% of its uses in all emoji). See examples in Fig. 5.14;

- *gem stone* 💎 (170 blends, 90.1% as modifier) uses “Replace part with object, Scale” in 84 blends. By considering all blends that use “Replace part with object”, the number rises to 112. We also identified a group of “Texture replacement” with 33 blends, which are more than half of all uses of “Texture replacement”;

- *balloon* 🎈 (164 blends, 81.2% as base) uses “Add contents, Rotation, Scale” in 71 blends. This number increases to 99 if we consider all blends that use “Add contents”;

- *turtle* 🐢 (162 blends, 85.3% as base) uses “Replace part with part, Scale” in 50 blends. This number increases to 95 if we consider all blends that use “Replace part with part”. “Add element” is also a frequent transformation, being used in 78 blends;

- *hot beverage* ☕ (151 blends, 84.8% as base) uses “Add contents, Scale” in 73 blends, which increases to 97 if we consider all blends that use “Add contents”;

- *heart suit* ❤️ (157 blends, 81.6% as modifier) uses “Replace part with object, Scale” in 75 blends. This number increases to 102 if we consider blends that use “Replace part with object”;

- *grinning face* 😀 (152 blends, 76.5% as modifier) uses “Replace part with part” as a unique transformation in 66 blends. This number rises to 110 if we consider blends that use the transformation among others. The pattern consists in using the smiling mouth (the most salient element) as replacement.

Despite the existence of these patterns, the transformations used are normally related to features of the emoji. For example, the *crown* 👑 is normally added on the top of the other emoji (*Add element*), resembling its normal use (see Fig. 5.14); blends with the *crystal ball* 🔮 that use *Object inside Object* take advantage of its usual purpose of showing something inside it (see Fig. 5.14); and in the case of *gem stone* 💎, its texture is used to turn other objects into gems. Moreover, when creating blends between two emoji with strong patterns, it is not obvious which strategy should be used – e.g. the blend produced for *crown* 👑

and *crystal ball* 🔮 is a crown inside the ball (pattern of the crystal ball, see Fig. 5.14) instead of a crown on top of the ball. These are emoji-specific patterns and cannot be used to draw general conclusions.

Two emoji from our selection, *heart suit* ♥️ and *grinning face* 😀, are particularly important to our study due to their similarity to other emoji. First, the *grinning face* 😀 shares resemblance to other emoji that can be grouped as "smiley faces". Most of the "smiley face" emoji also use the transformation *Replace part with part* but there is a higher tendency to either use *Add contents* or *Replace part with object*. While with *Replace part with part* there is a focus on the exchange of just one element (e.g. mouth), with *Add contents* and *Replace part with object* the blending process often concerns all the face elements. Moreover, in the case of "smiley faces", the blends produced using *Add contents* and *Replace part with object* are often similar, as the former consists in adding the inside elements (eyes, mouth, etc.) to the other object and the latter in replacing the yellow head circle with the other object.

Regarding the *heart suit* ♥️, the emoji is very similar to a set of 9 hearts in different colours – *black* 🖤, *blue* 💙, *brown* 🤎, *green* 💚, *red* ❤️, *orange* 🧡, *purple* 💜, *white* 🤍 and *yellow* 💛. Each of these hearts is used around 50 times except for the red heart, which is used fewer times (39) probably due to the high similarity with the *heart suit* ♥️. In all of the heart emoji, more than 50% of the blends (minimum of 64% in the *red heart* ❤️ and around 72% in the others) use only the transformation *Colour replacement*, which we consider a strong pattern. In this case, the most salient feature (the colour) is used as the ground for the transformation. It is interesting how there is a common strategy within the heart set but it differs from what happens with the *heart suit* ♥️ – we believe that since *Colour replacement* is already being used with the *red heart* ❤️, a different strategy had to be used with *heart suit*. On the other hand, *heart suit* belongs to a set of symbols related to card games. As the other symbols are not used in the *Emoji Kitchen* subset that we had access to, one may only wonder whether there would be similarity in terms of transformations used.

Based on these similarities between emoji, there is a question that remains: would it be possible to identify strategies used in emoji of certain types? To assess this, it is necessary to have a way of dividing emoji into groups. We will return to this topic in Chapter 17.6, in which we describe the use of an emoji categorisation to identify transformational patterns in groups of emoji.

5.4 DISCUSSION

As previously stated, the purpose of the study described in this chapter is to investigate Visual Blending in terms of the transformations used in the process of merging the initial inputs into a blend. As such, we moved away from common related research topics, such as visual

Table 5.3: *VisMet* and *Emoji Kitchen* structural analysis. Results of the analysis of the two image datasets in terms of structure, showing the total number of images (%T), the number of images analysed in the study (%A) and the percentage of each type of visual blend structure (F fusion, PJ pairwise juxtaposition, R replacement, RF replacing fusion and RJ replacing juxtaposition).

	%T	%A	F	PJ	R	RF	RJ
VisMet	353	125	29.6%	4.8%	16.8%	46.4%	2.4%
Emoji Kitchen	2243	2088	23.8%	20.45%	0.0%	55.75%	0.0%

metaphor, and focused on how visual blends are produced. To do this, we analysed two sets of images that contain visual blends – *VisMet* and *Emoji Kitchen*. The types of images of the two sets are very different. This difference can be observed in the reasons for excluding images: in *VisMet*, most exclusions were related to an excessive number of elements in the image; in *Emoji Kitchen*, the exclusions were due to an excess of elaboration or similarity to the initial input. Moreover, in *VisMet*, even though we consider that the analysed images show a visual blend, we are aware that in most cases an actual process of visual blending did not take place and no initial inputs exist. On the other hand, in *Emoji Kitchen* the blends are produced combining two emoji, resulting in a fairly simple visual representation. The *Emoji Kitchen* set seems to be based on the *Noto Emoji* set, enabling us to study the input of the blending process as well. For these reasons, the depth of analysis was greater with *Emoji Kitchen* than with *VisMet*.

We conducted an analysis mainly based on two levels: *structural* and *transformational*. For the structural level, we considered the taxonomy by Peterson (2018). In regards to the transformational level, existing taxonomies have issues when it comes to distinguishing different levels of analysis. We contribute to the field by presenting a novel taxonomy of visual blending transformations, which is a valuable resource for the analysis of visual blends and also for the implementation of visual blending systems.

Tables 5.3 and 5.4 show the results obtained with each dataset. As observed in Table 5.3, the two sets vary in terms of the type of structure of their blends. In both sets, RF is the most common type of structure, followed by F. An interesting result is how the percentage of PJ is much higher in *Emoji Kitchen* than in *VisMet*. Another result that shows the difference in nature of the two datasets is the absence of R in *Emoji Kitchen*. The absence of R is easy to justify: emoji are simple representations, normally focusing only on a single element, and, as such, there is not a sense of context, which is needed for Replacement (R). On the other hand, images from *VisMet* are much more complex and often include several elements, as previously described. A similar situation occurs with RJ, which requires the presence of several identical elements, one of which is replaced. This type of structure is also absent in the

Figure 5.15: Blending *cheese wedge* `U+1F9C0` and *fork and knife with plate* `U+1F37D`. Pairwise Juxtaposition using *Add element*, *Rotation*, *Scale*. Adapted from: *Emoji Kitchen* and *Noto Emoji*.

case of emoji. These results indicate that the type of structure depends on the type of input.

We again point out that there is a difference in *structural analysis* between considering the *input* as a *domain* – the perspective followed by Peterson (2018) – or a *visual object* – our perspective (e.g. a corkscrew and a wine bottle can be considered as two references to the domain *wine* but would be seen as two distinct visual objects). The definition of *pairwise juxtaposition* requires the presence of both inputs, which in the transformational perspective consists in showing the entire visual objects. In the analysis of *Emoji Kitchen*, we identified 29 cases in which a sort of *juxtaposition* is produced using one object and part of the other (see Fig. 5.15). In these cases, one could argue that one of the inputs (visual object) is not fully represented, which would go against the definition of PJ. As the number of these cases is reduced and our focus is on the transformations, we opted to label them as PJ. Despite this, we believe a new type of structure could be defined – "Partial Juxtaposition" – as these cases do not fit any of the remaining types.

From our analysis of *Emoji Kitchen*, we were able to draw some conclusions related to the use of emoji input to produce visual blends:

- Certain emoji are more prone to be base and modifier;
- Certain emoji use certain transformations more than others in their blends;
- Certain groups of objects have transformational patterns (e.g. the "hearts" and "smiley faces").

At the beginning of the chapter, we identified two questions that we aimed to address with this study. The first question asked whether it would be possible to identify transformations used in the production of visual blends and derive general transformational patterns. By analysing both sets, we were able to identify the most used transformations (see Table 5.4). In the case of *Emoji Kitchen*, we found transformation patterns in certain emoji. As our main goal is to produce general knowledge that can be used when implementing a visual blending system, emoji-specific transformations are not very useful to draw general conclusions. To address this issue, an approach to identify general patterns would be needed. In Chapter 17, we return to this topic and we use an emoji categorisation oriented towards visual blending to identify transformational patterns.

By comparing the results of the analysis of both sets, we can see that the type of transformations used varies. An interesting result concerns the values of *Add element* and *Replace part with object*, which are high in both sets. A considerable difference can be seen in the percentage of *Scale*, which obtained 0 % in *VisMet* and 77% in *Emoji Kitchen*. This difference is easy to understand if we consider that scale is difficult to identify without the initial input, hence the absence in *VisMet*, and it is often used for placing purposes in *Emoji Kitchen*, hence the high value.

In general, we consider that using a transformational approach when analysing blends may lead to more concrete conclusions in how visual blends are produced and may allow for a more objective comparison between images. However, we are also aware that a given blend can be produced by using different combinations of transformations, which makes it more complex to analyse images.

The second question asked whether our findings could be used in a computational system to general visual blends. The answer to this second question is complex as it raises several other questions.

First, our findings concern several stages of the blending process. We were able to identify strategies that are used to map elements and transformations that are often employed. In terms of aspects that are used for mapping, our results are in accordance with previous studies, especially in regards to *perceptual similarity*. We were able to identify five types of *alignment* – *colour*, *shape*, *proportion*, *texture* and *conceptual* – that are used to establish the mappings upon which the blend will be based. This is especially obvious in the case of RF in *Emoji Kitchen*, in which we identified alignment in 1085 out of the 1164 RF blends. Alignment can be employed in computational systems to identify possible exchange candidates, as it is already the case of *shape similarity* used by Chilton, Petridis, and Agrawala (2019). In this case, the comparison is done in terms of *perceptual similarity*, which can be implemented by using several techniques (e.g. *colour comparison*). Conceptual similarity, on the other hand, is more difficult to assess as it is not *visual* but *semantic*, which requires that the system has access to information about each visual element. Such information is not usually available and systems may have to resort to other approaches – e.g. in some cases, conceptually similar elements are located in a similar place, provided that style is maintained (e.g. face emoji). In any case, in visual blending it is advantageous to use perceptually similar elements, contributing to higher coherence and integration, and improving the recall of replaced elements (as shown by the studies described in Section 4.2.3). All in all, if producing a blend based on a blind exchange of parts we can focus on similarity and do mappings based on it.

In regards to transformations, both analysis were important to assess the most used transformations. Despite this, *Emoji Kitchen* comes closer

Table 5.4: *VisMet* and *Emoji Kitchen* transformational analysis. Results of the two image datasets for each type of transformation. The table shows the calculated percentage in terms of the total number of analysed blends (%) and counts for each transformation – the total blend count (#), the number of blends that uniquely use the transformation (U) and the count for each type of structure (F fusion, PJ pairwise juxtaposition, R replacement, RF replacing fusion and RJ replacing juxtaposition).

	VisMet								*Emoji Kitchen*							
	%	#	U	F	PJ	R	RF	RJ	%	#	U	F	PJ	RF	R	RJ
Add contents	2.4%	3	2	3	0	0	0	0	21.7%	453	1	19	0	434	0	0
Add element	22.4%	28	14	20	6	1	1	0	26.5%	553	6	75	360	118	0	0
Colour replacement	2.4%	3	0	1	0	0	1	1	29.5%	617	322	407	11	199	0	0
Flip	0.0%	0	0	0	0	0	0	0	0.7%	14	0	1	12	1	0	0
Keep colour	3.2%	4	0	2	0	1	1	0	1.2%	25	0	1	0	24	0	0
Keep contents	2.4%	3	0	0	0	0	3	0	0.0%	1	0	1	0	0	0	0
Keep shape	1.6%	2	0	0	0	0	2	0	0.2%	5	0	1	0	4	0	0
Keep texture	4.0%	5	0	1	0	1	3	0	0.8%	17	0	3	0	14	0	0
Merge element	7.2%	9	7	9	0	0	0	0	0.2%	5	0	5	0	0	0	0
Object inside Object	1.6%	2	0	1	1	0	0	0	3.2%	67	0	0	67	0	0	0
Object with missing part	0.8%	1	0	1	0	0	0	0	0.1%	2	0	1	0	1	0	0
Replace part with object	40.0%	50	36	0	0	18	29	3	24.4%	510	10	12	0	498	0	0
Replace part with part	8.0%	10	9	1	0	2	7	0	10.5%	220	67	16	0	204	0	0
Rotation	2.4%	3	0	0	1	1	1	0	14.8%	309	1	9	160	140	0	0
Shape replacement	11.2%	14	11	4	0	1	9	0	1.8%	37	5	28	0	9	0	0
Scale	0.0%	0	0	0	0	0	0	0	77.0%	1607	4	116	426	1065	0	0
Texture from separate objects	7.2%	9	7	2	0	0	7	0	0.0%	1	0	1	0	0	0	0
Texture replacement	11.2%	14	7	7	0	0	7	0	3.1%	64	29	61	0	3	0	0

to an implementation scenario as we have both inputs and final blends. Overall, we consider that the majority of the identified transformations are easy to include in a system for visual blending. Nonetheless, there are cases of blends that would be difficult to produce in an automatic or even semi-automatic system – e.g. the blend between *face with head-bandage* 🤕 (`U+1F915`) and *hot beverage* ☕ (`U+2615`), which shows a broken cup (Fig. 5.10). This blend is resultant from an elaboration by the designer, who is not limited to the use of a small set of transformations and has enough freedom to include elements that are not present in the initial input. In addition to this, certain blends go beyond simple transformations, for example: in blends with *hole* emoji 🕳 (`U+1F573`), a partial mask is often used to give the impression that an element en-

(a) *angry face* (b) *pouting face*

Figure 5.16: Comparison between *angry face* emoji (`U+1F620`) and *pouting face* emoji (`U+1F621`).

ters the hole (Fig. 5.17); and in blends with the *crystal ball* 🔮 (`U+1F52E`), the added elements are positioned behind the ball element, which is then made semi-transparent to give the impression that the element is inside it (Fig. 5.18). In these two cases, there are emoji-specific transformations that extend the ones included in the taxonomy.

Figure 5.17: Blend between *hole* (`U+1F573`) and *newspaper* (`U+1F4F0`). Source: *Emoji Kitchen*

Figure 5.18: Blend between *heart suit* (`U+2665`) and *crystal ball* (`U+1F52E`). Source: *Emoji Kitchen*

Second, the aspects that influence the blending process go beyond mere transformational patterns. This fact became evident during our analysis of the visual blends. An aspect that has high relevance when choosing key elements for the blending process is *saliency*. Saliency can be interpreted as the importance a given element or property has in a visual representation. In the case of emoji, one could even say that saliency is closely related to the idea of *prototypicality* (see Section 2.2.1), as emoji can in a way be seen as simplified visual representations of *core concepts*. An example of this can be observed when comparing the *angry face* emoji 😠 (`U+1F620`) with the *pouting face* emoji 😡 (`U+1F621`), see Fig. 5.16. The *angry face* consists of a simple face, with eyes and eyebrows scrunched downward, and a frowning mouth. The *pouting face*, which is associated with hate or range, contains exactly the same elements but with a red face instead of a yellow one. In this case, it is easy to identify the red colour as being prototypical of "pouting". When producing a visual blend with the *pouting face* 😡 (`U+1F621`) emoji, a *Colour replacement* transformation would likely be used to assign the red colour to the blend. Other examples of salient features have been mentioned throughout the chapter, e.g. face elements (e.g. mouth) in smileys or different colours in coloured hearts. Concerning face elements, one can see the relevance of salient features when analysing four blends resultant from the combination of two identical face emoji: the blends consist in exaggerating the most salient element (e.g. mouth or eyes) by using a *Scale* transformation (an example is shown in Fig. 5.19). In practice, for a given visual representation one should identify the most salient elements, as these are likely to have an important role in a process of visual blending. By using a dataset of images of the same style, as is the case of emoji, a possible solution is to compare all images to identify the most distinct features of each image. In this sense, emoji seem to be suitable to be used in a system for visual blending, due to their unique characteristics: *simplicity* and

Figure 5.19: Blending between two *grimacing face* `U+1F62C`. Adapted from: *Emoji Kitchen* and *Noto Emoji*.

consistency of style. Although these characteristics may be present in image datasets, salient features may not be easy to identify. Take, for example, an image of a *cactus*, which shows a repetition of spines. In this case, the repetition of spines is a *prototypical feature* and it would likely be used as a texture in a visual blending process. Here, a comparison with other images would not be sufficient and a within image assessment of similar elements is also necessary. Another example is shown in Fig. 5.20, in which the salient feature of *face without mouth* 😶 (`U+1F636`) is not an element but the absence of one (mouth). In this case, this absence is used in the blending processing through the removal of the eyes of *grinning face* 😀 (`U+1F600`).

Another important aspect concerns the number of performed exchanges in the visual blend. Consider, for example, the blend of a *person* with a *pig*. There are multiple solutions to this blend, some more complex than others. It can be as simple as replacing the *nose* of the person with the pig's or be more complex by adding or exchanging more elements (e.g. a tail) or changing their properties (e.g. colour). One question that remains is: how should the system decide which solution is best? Again, the answer to this question involves more than simple transformations.

When producing a visual blend there is often an intended message and deciding on which mappings to make and which transformations to use is highly dependent on the meaning that is to be encoded. One of the biggest issues of *Emoji Kitchen* is that the blends are "forced" combinations, produced without any given meaning as a goal. On the one hand, the analysis of *Emoji Kitchen* blends is good to identify possible "default" solutions, which may provide clues on how to produce the most visually interesting blends. On the other hand, it does not allow for conclusions related to meaning. An open question is how to choose the set of transformations to be made when producing a blend – such is not clear in the case of *Emoji Kitchen*. Although certain emoji seem to have more importance in this decision and a given type of transformation may be more likely to occur, the answer is not straightforward. For example, when blending a *crown* and a *crystal ball*, which have clear transformational patterns according to our results, there are at least two obvious solutions: the crown on top of the ball and the crown inside the ball. Despite this, the solution to this blend is not trivial. The transformation choice is highly dependent on the intended meaning. In the case of the solution "crown inside the crystal ball", it may lead

Figure 5.20: Blending between *face without mouth* `U+1F636` and *grinning face* `U+1F600`. Adapted from: *Emoji Kitchen* and *Noto Emoji*.

to the interpretation of foreseeing a future related to royalty. The interpretation of the other solution, "crown on top of the crystal ball", may be considerably different. This shows that different combinations lead to different meanings and that certain visual elements have specific meanings associated, which may have a great impact on the blend. Despite being an important aspect, the study of meaning is beyond the scope of the study described in this chapter.

5.5 SUMMARY

In this chapter, we proposed a taxonomy of transformations that can be used in visual blending. Then, we described a study based on the analysis of visual blends from two different datasets (*VisMet* and *Emoji Kitchen*), using the proposed taxonomy. With this study, we aim to identify the most common transformations used in the visual blends of the two datasets.

In summary, the findings from the study are useful for computational approaches to visual blending and can be used at the different stages of the blending process. We have also described how several aspects have an impact on the production of visual blends: *alignment*, *salient features*, *element properties* and *meaning*. The main contributions of this chapter are: (i) the proposal of a new visual blend transformation taxonomy; and (ii) the description of two studies with visual blends, in which we apply the proposed taxonomy and identify the transformations most used in the blends.

6 A PIG, AN ANGEL AND A CACTUS WALKED INTO A BLENDER

In the previous chapters, we have introduced visual blending as a process that can be used for the visual representation of concepts. In this chapter, we describe our first exploration with visual blending, which we consider the first case study of the thesis. We present a descriptive approach for the automatic generation of visual blends. The implemented system, the *Blender*, is composed of two components: the *Mapper* and the *Visual Blender*. We conduct experiments with three initial input concepts *pig*, *angel* and *cactus*. The performance of the system is analysed by comparing the produced blends with user-drawn ones and also assessing how users perceive the blends.

6.1 CONTEXT

As previously mentioned, one of the earliest works to computationally produce visual blends is the *Boat-House* Visual Blending Experience (Pereira and Cardoso, 2002). In their work, Pereira and Cardoso (2002) use a Logo-like programming language to produce visual representations. These representations can be considered visualisations of blends produced at the conceptual level. One interesting aspect of their work is the style used in the visual representations – simple yet containing the most representative elements of each concept (e.g. wall, roof, door and window in the case of *house*). This approach is aligned with a process of reduction to essential parts, which can be seen as the base for a prototypical representation of some concrete objects (as described in Section 2.2). By using a simple representation style, e.g. one that consists of basic shapes, it is possible to reduce the concept to its simplest form, while maintaining its most important features and thus, hopefully, capturing its essence. This approach is similar to the simplification process used by Picasso in *The Bull* (see Chapter 2 for a more thorough description).

This approach based on simplification makes it possible to experiment with visual and conceptual blending, in a context of reduced complexity. Such would be difficult when using, for example, photorealistic input images.

The work by Engelhardt (2002) (explained in Section 2.4) can also be used as support to this approach. Engelhardt (2002) proposes that a *composite graphic object* consists of: a *graphic space* (occupied by the *object*); a set of *graphic objects* (which may also be *composite graphic objects*); and a set of *graphic relations* (which may be *object-to-space* and/or *object-to-object*). First, this type of structure may bring advantages to the process of visual blending, by facilitating the exchange of objects. Second, the relations between graphic objects may be aligned with relations among concepts in mental spaces, thus creating a connection between the conceptual and visual layers. The approach described in this chapter is based on these two points.

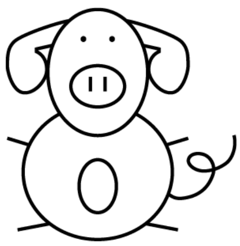
6.2 IMPLEMENTATION

Our approach is based on an *alignment* between the *conceptual level* and the *visual level*. The goal is to implement a hybrid system, in which the *blending process* starts at the conceptual level but only ends at the visual one. In this sense, the visual representations are not merely a visualisation of blends previously produced at the conceptual level but the component that controls the visual output also has a role in the process of producing the blends.

As such, the system receives a set of *concepts* and produces *blends* from them, relying on two types of input for each concept: (i) an *input mental space* in the form of a *semantic network* and (ii) *visual representations*. In order to achieve an alignment between the two levels, at least part of the elements of the mental space should be represented in the visual representation (e.g. *head* is represented as a node in the semantic network and as a circle in the visual representation).

Having the organisation of *mental spaces* as an inspiration, we consider a *visual representation* (referred to as "representation") as a group of several parts/elements. This way, the representation ceases to be a whole and starts to be seen as parts related to each other. As we wish to produce visual results, these relations have a visual descriptive nature (i.e. the nature of the relation between two elements is either related to their relative position or to their visual qualities). Our goal is that these visual relations can be matched with relations between *nodes* in the *mental spaces* and ultimately used to produce *analogies* between two given concepts. This results in the generation of visual blends, guided by the produced analogies (based on conceptual relations) and evaluated using the criteria imposed by the visual relations among parts of the input visual representations (base representations).

The approach is centred on the idea that the construction of a visual representation for a given concept can be approached in a structured way, following the proposal by Engelhardt (2002). Each visual representation is associated with a list of descriptive relations (e.g.: *part A* below *part B*), which describes how the representation is constructed (see example in Fig. 6.1). Due to this, a visual blend between two input visual representations is not simply a replacement of parts but its quality is assessed based on the number of relations that are respected. This gives much more flexibility to the construction of visual represen-



inside([left_eye,right_eye],head)
upperPart([left_eye,right_eye],head)
left(left_eye,right_eye)
inside(nose,head)
lowerPart(nose,head)
touch([left_ear,right_ear],head)
upperPart([left_ear,right_ear],head)
leftPart(left_ear,head)
rightPart(right_ear,head)
behind(body,head)
upperPart(head,body)
touch(head,body)
middle(shapebelly,body)
touch([right_arm,left_arm,right_leg,left_leg,tail],body)
rightPart([right_arm,right_leg,tail],body)
leftPart([left_arm,left_leg],body)
upperPart([right_arm,left_arm],body)
lowerPart([right_leg,left_leg],body)
surroundedBy(tail,[right_arm,right_leg])

Figure 6.1: Visual representation and visual relations

tations by presenting a version of it and also allowing the generation of similar ones if needed.

In summary, the goal is to implement a system that receives input *visual representations* and *semantic networks* for a set of *concepts* and produces *visual blends* based on them. In this section, we first describe the process of collecting the data necessary for implementing the system and then we describe the implementation of the system itself.

6.2.1 *Collecting data*

An initial phase of data collection took place to gather the necessary materials to implement the system. The following materials were collected: *concepts* to use, *visual relations* and *visual representations* for each concept.

6.2.1.1 *Concepts*

The initial phase of the project consisted in a process of data collection. First, a list of possible concepts was produced by collecting concepts used in research conducted related to conceptual combination and conceptual blending. The collection process was not extensive as our goal was to only select a few concepts. Some examples of collected concepts are presented in Table 6.1. From this list, we made a final selection of concepts to be used in our experiment. For our selection, we focused on noun-noun concrete concepts that could have literal visual representation,[1] thus excluding concepts such as *pet fish* and *black bird*. Also, we decided to only address concepts of the natural kind,[2] excluding artefact ones (e.g. *gun*). Three concepts were selected based on their characteristics: *angel* (human-like), *pig* (animal) and *cactus* (plant).

1 Pereira (2004, p. 195) presents a list of conceptual blending examples but most of them do not have a clear visual representation.
2 Costello and Keane (1997) present a study in which noun-noun compounds are compared based on their composition, i.e. using artefact or natural terms, showing that combinations involving artefact terms are more polysemous.

Table 6.1: Examples of collected concepts. Sources: [1] (Costello and Keane, 1997), [2] (Costello and Keane, 2000), [3] (Keane and Costello, 2001),[4] (Costello and Keane, 2001), [5] (Costello, 2002), [6] (Pereira, 2004), [7] (Coulson, 2006).

Concept	Sources
pet fish	[1][2][3][6][7]
cactus fish	[1][2][3][4]
bumblebee moth	[2][3][4][5]
...	
angel pig	[1][3][6]
apartment dog	[3][4][5]
kangaroo monkey	[3][4][5]
...	
elephant gun	[1]
head hat	[2]

6.2.1.2 *Visual Relations*

As previously explained, the visual representation used as input, as well as those produced by the system are structured in *layers* and accompanied by a description of the existing connections among their elements, using *visual relations*. To achieve this, we produced a list of visual relations that would be used to structure the visual representations. This list is the result of an analysis of existing work on visual relations, which was described in Section 2.4.2.

The produced list contains relations of the three most used types of *spatial-based relations* (*directional*, *topological* and of *distance*), as identified by Wang et al. (2008), and also includes *attribute-based relations*. In this latter group, *colour*- or *texture-based* relations could also have been included but we chose not to add them as our intention was to only use black and white visual representations. The produced list is presented in Fig. 6.2.

6.2.1.3 *Visual Representations*

Having chosen the concepts to use in the experimentation and defined the visual relations to use, we still needed to obtain visual representations for each concept. The initial idea consisted in only having a visual representation for each concept. However, a given concept has several possible representations (e.g. there are several possible ways of visually representing the concept *car*), which means that only using one would increase the limitations of the system. In order to increase the variety of results, we decided to use several versions for each concept.

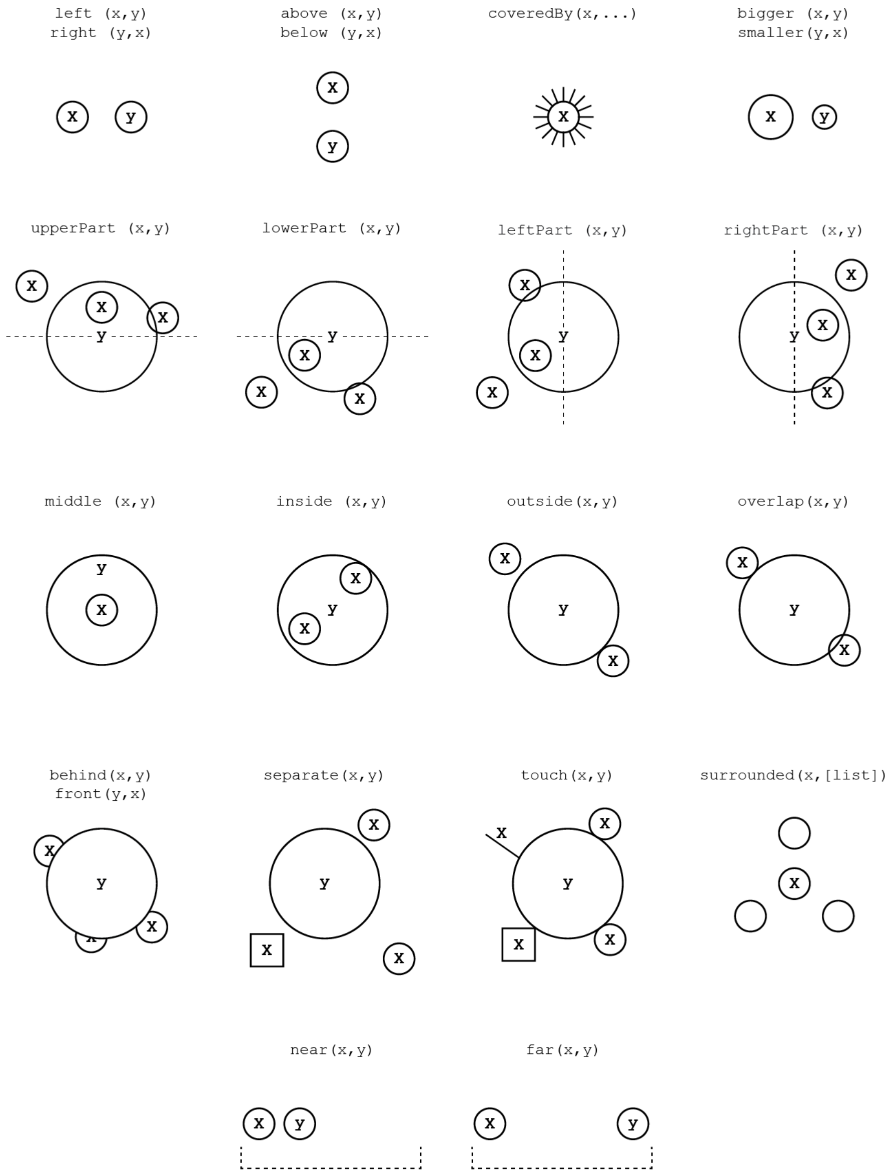


Figure 6.2: List of visual relations. The list includes *spatial-based relations* of three types – *directional* (e.g. *left*, *above*, etc.), *topological* (*overlap*, *touch*, etc.) and of *distance* (*near* and *far*) – and *attribute-based* relations (*bigger* and *smaller*).

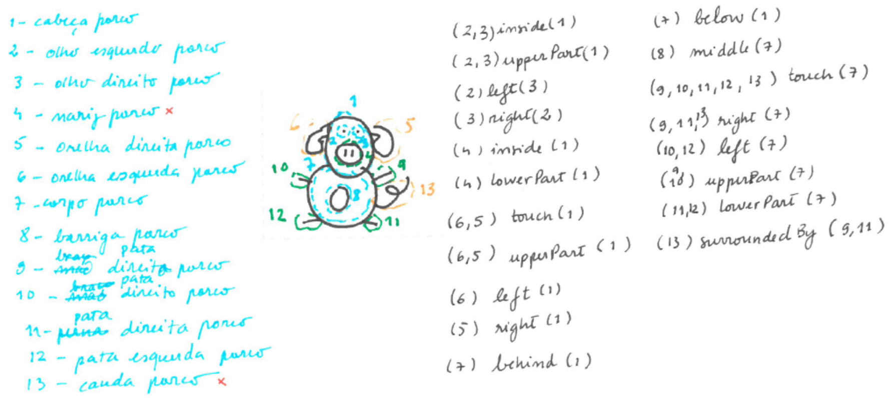


Figure 6.3: Identification of parts and visual relations for a visual representation of *pig* by a user.

Each visual representation can be different (varying in terms of *style*, *complexity*, *number of characteristics* and even chosen *perspective*) and thus also have a different set of visual relations among the parts.

As such, the goal of this phase was to collect visual representations for the selected concepts. To do this, we designed a user enquiry, which was composed of five tasks:

T1 Collection of visual representations for the selected concepts;

T2 Identification of the representational elements;

T3 Description of the relations among the identified elements;

T4 Identification of the prototypical elements – i.e. the element(s) that most identify a given concept. For instance, most participants considered *nose* and *tail* as the prototypical elements of *pig*;

T5 Collection of visual blends for the selected concepts.

The enquiry was conducted with nine participants, who were asked to complete the five tasks. In the first task (T1), the participants were asked to draw a representation for each concept avoiding unnecessary complexity but still representing the most important elements of the concept. In order to achieve intelligible and relatively simple representations, the participants were suggested to use *primitives* such as *lines*, *ellipses*, *triangles* and *quadrilaterals* as the basis for their drawings. After completing the first version, a second one was requested. The reason for the two versions was to promote diversity.

In the second task (T2), the participants identified the drawn elements using their own terms (for example, for the concept *angel* some of the identified elements were *head*, *halo*, *legs*), see example in Fig. 6.3.

After completing the previous task, the participants were asked to identify the relations among elements that they considered as being essential (T3), see example in Fig. 6.3. These relations were not only

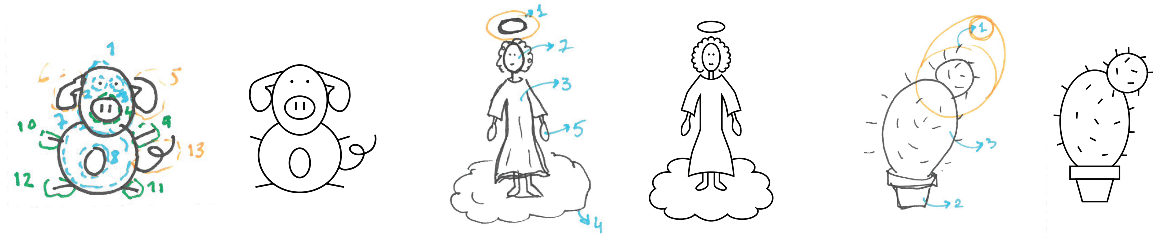


Figure 6.4: Conversion to Scalable Vector Graphics (SVG). For each concept, on the left is a representation drawn with the elements identified and on the right is the result of the conversion into SVG.

related to the conceptual space but also (and mostly) to the representation. In order to help the participants, a list of visual relations was provided (see Fig. 6.2). Despite being told that the list was only an example and not to be seen as closed, all the participants used the relations provided – promoting an alignment among participants.

The identified relations are dependent on the author's interpretation of the concept, which can be divided into two levels. The first level is related to how the author interprets the connections among the concepts and their parts at a conceptual level (for example *car*, *wheel* or *trunk*). The second level is related to the visual representation being considered: different visual representations may have different relations among the same parts (this can be caused, for example, by the change of *perspective* or *style*) – e.g. the different positioning of the head in the two *pig* representations in Fig. 6.5.

Task four (T4) consisted in identifying the *prototypical parts* of the representations (see elements marked with "x" in Fig. 6.3) – i.e. the parts which most identify the concept (see Section 2.2.1 for a description of the *Prototype Theory*). The identification of these parts is useful for analysing the results obtained with the system.

In the last task of the enquiry (T5), the participants were asked to draw visual representations for the blends between the three concepts. As a blend between two concepts can be interpreted and posteriorly represented in different ways (e.g. just at a naming level a blend between *pig* and *cactus* can be differently interpreted depending on its name being *pig-cactus* or *cactus-pig*). For this reason, the participants were asked to draw one or more visual representations for the blend. These visual representations were later used for comparing with the results obtained with the *Visual Blender*.

After the conduction of the enquiry, the data was treated in order to be used by the *Visual Blender*. Firstly, the representations collected for each of the concepts were converted into SVG format (see Fig. 6.4) in which each part was placed in a separate layer, and prepared to be used as base visual representations (see Fig. 6.5) for the *Visual Blender*, using layer naming according to the data collected for each representation – each layer was named after its identified part. In addition to this, the

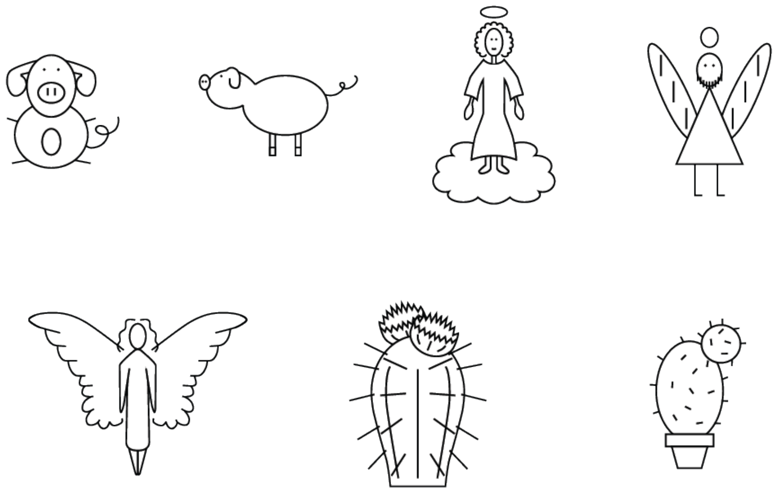


Figure 6.5: Representations used as a base.

relations among parts were formatted to be used as input together with their corresponding visual representation (Fig. 6.1).

6.2.2 *Implementing The Visual Blender*

In the previous sections, we described the process of collecting the data necessary to produce visual blends. In this section, we describe the implementation of the system.

As already mentioned, the *Blender* has two different components: the *Mapper* and the *Visual Blender* (see Fig. 6.6). The *Mapper* receives two *input spaces* (represented as 1 in Fig. 6.6), one referring to *concept A* and the other one to *concept B*. It produces *analogies* (3 in Fig. 6.6) that are afterwards used by the *Visual Blender* component. The *Visual Blender* also receives *visual representations* and a corresponding *list of relations* among parts (2 in Fig. 6.6) that are used for producing the *visual blends* (4 in Fig. 6.6).

Our work is focused on the *Visual Blender* component. The *Mapper* component is only briefly described (see Section 6.2.2.1), as it is not a contribution of this thesis.

6.2.2.1 *The Mapper*

In Conceptual Blending (CB) theory, after the selection of *input spaces*, the subsequent step is to perform a partial matching between elements of the given mental spaces. This can be seen as establishing an *analogy* between the two inputs. The input spaces are in the form of semantic maps composed of N_c *concepts* and N_t *triples*, with $N_t, N_c \in \mathbb{N}$. The triples are in the form $< concept_i, relation, concept_j >$. $concept_i$ and $concept_j$ correspond to vertices in a graph, which are connected by the directed edge labeled as *relation* from $concept_i$ to $concept_j$.

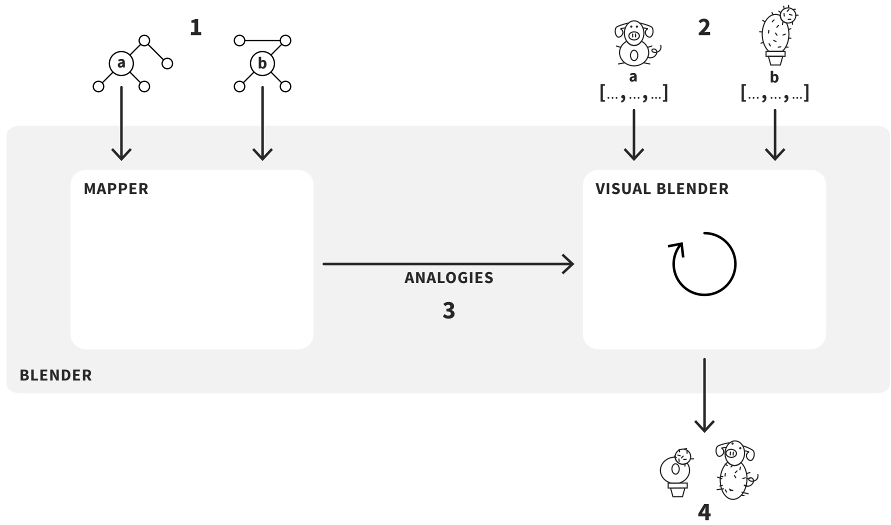


Figure 6.6: Structure of the implemented *Blender*. The *Blender* consists of a *Mapper* and a *Visual Blender*. The figure also shows the input spaces (1), the visual representations and list of relations (2), the produced analogies (3) and the produced blends (4).

The *Mapper* iterates through all possible root mappings, each composed of two distinct concepts taken from the input spaces. This means that there is a total of $\binom{N_c}{2}$ iterations. Then, the algorithm extracts two isomorphic *sub-graphs* from the larger input space. These sub-graphs are split into two *sets of vertices* A (left) and B (right). The *structural isomorphism* is defined by the sequence of relation types (PW, ISA, etc.) found in both sub-graphs.

Starting at the root mapping defined by two concepts (left and right), the isomorphic sub-graphs are extracted from the larger semantic structure (the input spaces) by executing two synchronised expansions of nearby concepts at increasingly depths. The first expansion starts from the left concept and the second from the right concept. The left expansion is done recursively in the form of a *depth-first expansion* and the right as a *breadth-first expansion*. The synchronisation is controlled by two mechanisms:

1. the depth of the expansion, which is related to the number of relations reached by each expansion, starting at either concept from the root mapping;

2. the label used for selecting the same relation to be expanded next in both sub-graphs.

Both left (depth) and right (breadth) expansions are always synchronised at the same level of deepness (first mechanism above).

While expanding, the algorithm stores additional associations between each matched relations and the corresponding concept which was reached through that relation. In reality, what is likely to happen is the occurrence of a multitude of isomorphisms. In that case, the algorithm will store various mappings from any given concept to multiple different concepts, as long as the same concepts were reached from a previous concept with the same relation. In the end, each isomorphism and corresponding set of concept mappings gives rise to an *analogy*. The output of the *Mapper* component is a *list of analogies* with the greatest number of mappings.

We refer the reader to Gonçalves, Martins, and Cardoso (2018) for a more detailed description.

6.2.2.2 *Generating the blends: construction and relations*

The *Visual Blender* component uses structured base visual representations (of the input concepts) along with their set of relations among parts to produce visual blends based on *analogies* (set of mappings) produced by the *Mapper* component.

The way of structuring the representations is based on the *Syntactic Decomposition of Graphic Representations* proposed by Engelhardt (2002). As such, we consider that each visual representation is composed of several *graphical objects* or elements. The objects store the following attributes: *name*, *shape*, *position* relative to the father-object (which has the object in the set of graphic objects), the *set of relations* to other objects and the *set of child-objects*. By having such a structure, the complexity of blending two base representations is reduced, as it facilitates object exchange and recursive changing (by moving an object, the child-objects are also easily moved).

A relation between two objects consists of: the object A, the object B and the type of relation (*above, lowerPart, inside, ...*) – e.g. *eye* (*A*) <u>inside</u> *head* (*B*), as shown in Fig. 6.1.

6.2.2.3 *Generating the blends: visual blending*

The *Visual Blender* receives the analogies between two given concepts produced by the *Mapper* component and the blending step occurs during the production of the visual representation – differently from what happens in *The Boat-House Visual Blending Experience* (Pereira and Cardoso, 2002), in which the blends are merely interpreted at the visual representation level.

The part of the blending process that occurs at the *Visual Blender* produces visual representations as output and consists of five steps:

S1 An analogy is selected from the set of analogies provided by the *Mapper*.

S2 One of the concepts (either *A* or *B*) is chosen as a *base* (as an example, consider *A* as the chosen one).

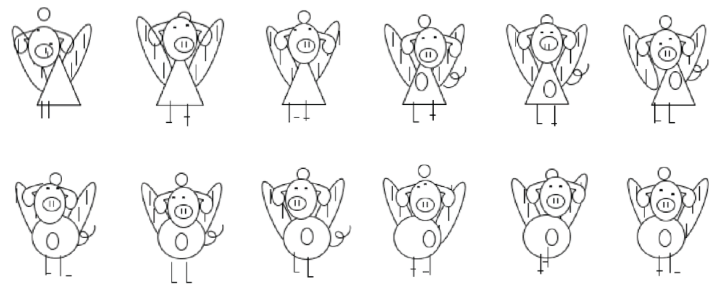


Figure 6.7: Different angel-pigs produced using the same or similar rules, the produced results are still quite diverse.

S3 A visual representation (rA) is chosen for the concept A and a visual representation (rB) is chosen for the concept B.

S4 Parts of rA are replaced by parts of rB based on the analogy. For each mapping of the analogy – consider for example *leg* of A corresponds to *arm* of B – the following steps occur:

S4.1 The parts from rA that correspond to the element in the mapping (e.g. *leg*) are searched using the names of the objects. In the current example, the parts found could be *left_leg* (*left_* is a prefix), *right_leg_1* (*right_* is a prefix and *_1* a suffix) or even *leftfront_leg*.

S4.2 For each of the found parts in S4.1, a part that matches is searched in rB using the names of the objects. This search firstly focus on objects that match the full name, including the prefix and suffix (e.g. *right_arm_1*) and, if none is found, the search is extended to parts with only the mapping name (e.g. *arm*). It avoids plural objects (e.g. *arms*). If no part is found, it proceeds to step S4.4.

S4.3 The found part (pA) of rA is replaced by the matching part (pB) of rB, updating the relative positions of pB and its child-objects, and relations (i.e. relations that used to belong to pA now point to pB).

S4.4 A process of Composition occurs (see examples in Fig. 6.7 – the *tail* and the *belly / round shape* in the triangular *body* are obtained using composition). For each of the matching parts from rB (even if the replacement does not occur) a search is done for parts from rB that have a relation with pB (for example, a found part could be *hand*). It only accepts a part if rA does not have a part with the same name and if the analogy used does not have a mapping for it. If a found part matches these criteria, a composition can occur by copying the part to rA (in our example, depending on either the replacement in Step S4.3 occurred or not, rA would have either *hand* related to *arm* or to *leg*, respectively).

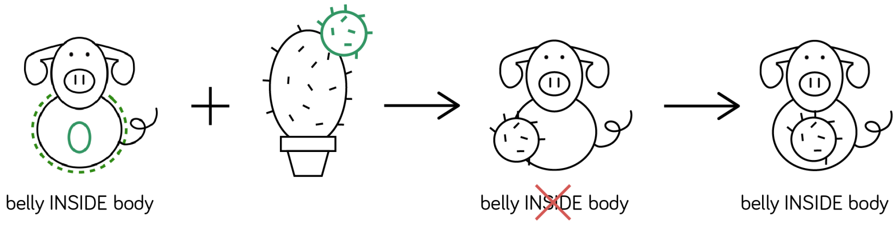


Figure 6.8: Visual relations being respected

s5 The *rA* resulting from the previous steps is checked for inconsistencies (relative positioning and obsolete relations), which can happen if an object ceases to exist due to a replacement.

After generating a representation, the similarity to the base representations (*rA* and *rB*) is assessed to avoid producing visually equal representations. This assessment is done by using a Root Mean Square Error (RMSE) measure that checks the *pixel-by-pixel* similarity.

6.2.2.4 *Evolutionary Engine*

The main goal of the *Visual Blender* component is to produce and evolve valid visual blends based on the analogies produced by the *Mapper*. To achieve this and promote diversity while respecting each analogy, an evolutionary engine was implemented. This engine is based on a Evolutionary Algorithm (EA) that uses several populations (each corresponding to a different analogy), in which each individual is a blend.

In order to guide evolution, we adopt a fitness function that assesses how well the existing relations are respected. Some of the relations, e.g. the relation *above*, have a binary assessment – either 0 when the relation is not respected, or 1 when it is respected (see Fig. 6.8). Others yield a value between 0 and 1 depending on how respected it is – e.g. the relation *inside* is based on the number of points of an object that are inside another object, obtaining its value as follows:

$$inside\_value = \frac{\#PointsInside}{\#Points}. \tag{6.1}$$

The fitness function for a given visual blend b is as follows:

$$f(b) = \frac{\sum_{i=1}^{\#R(b)} v(r_i(b))}{\#R(b)}, \tag{6.2}$$

where $\#R(b)$ denotes the number of relations present in b and v is the function with values in $[0,1]$ that indicates how much a relation r is respected, from not respected at all (0) to fully respected (1).

The evolutionary engine includes five tasks that are performed in each generation for each population:

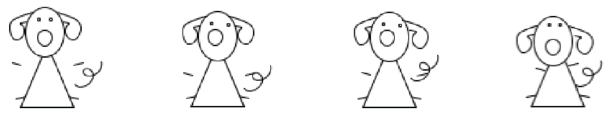


Figure 6.9: Evolution of a blend: the *legs* and *tail* come closer to the *body*, guided by the fitness function.

T1 Produce more individuals when the population size is below the maximum size;

T2 Store the best individual to avoid losing it (*elitism*);

T3 Mutate the individuals of the population. For each individual, each object can be mutated by changing its position. This change also affects its child-objects;

T4 Recombine the individuals: the parents are chosen using tournament selection (with size 2) and a *N-point crossover* is used to produce the children. In order to avoid the generation of invalid individuals, the crossover only occurs between chromosomes (objects) with the same name (e.g. a *head* is only exchanged with a *head*). If this rule was not used, it would lead to the production of descendants that would not respect the analogy followed by the population;

T5 Removal of identical individuals in order to increase variability.

In the experiments reported here, the mutation probability is set to 0.05, per gene, and the recombination probability to 0.2, per individual. These values were established empirically in preliminary runs.

6.3 GENERAL ANALYSIS

In this section, we present and discuss the results obtained. We begin with a general analysis. Then, we analyse the resulting visual representations comparing them with the data collected in the initial enquiry.

Overall, the analysis of the results indicates that the implemented blender is able to produce sets of blends with high variability (see Fig. 6.7 for an example of the results obtained for the same analogy and the same relations) and unexpected features while respecting the analogy. The evolutionary engine is capable of evolving the blends towards a higher number of satisfied relations. This is verifiable in numerical terms, through the analysis of the evolution of fitness, and also through the visual assessment of the results. Figure 6.9 illustrates the evolution of a blend: the *legs* and *tail* are iteratively moved towards the *body* in order to increase the degree of satisfaction of the relations.

We can also observe that the system tends to produce blends in which few parts are exchanged between concepts. This can be explained as

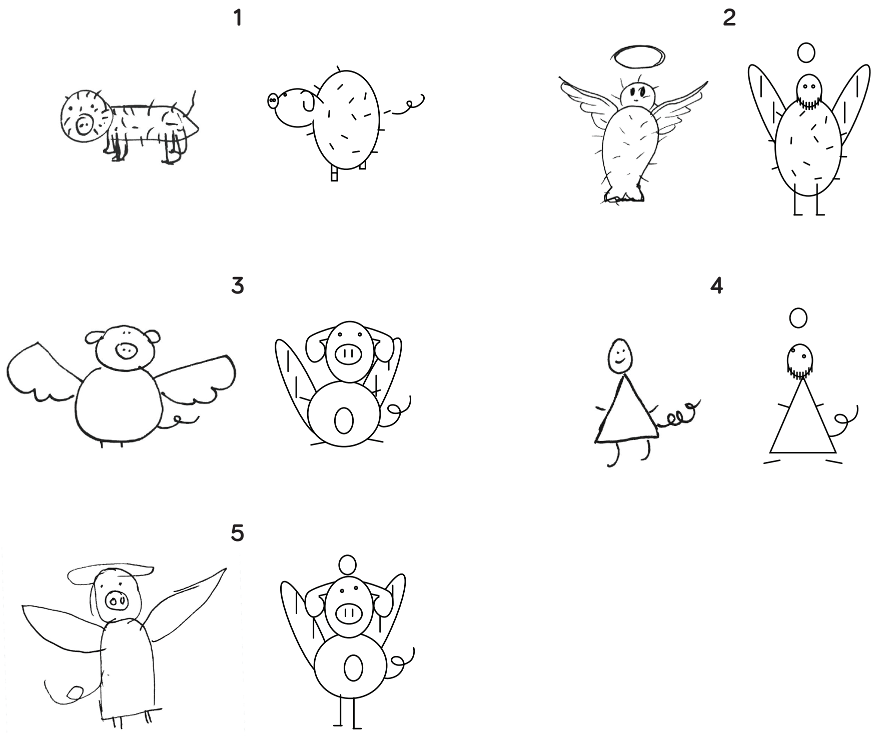


Figure 6.10: Comparison between hand-drawn blends and blends generated by the implemented *Blender*, organised by groups: group 1 corresponds to *pig-cactus blends*; 2 corresponds to *angel-cactus*; groups 3-5 correspond to *pig-angel* (the figure on the left of each group is the hand-drawn blend).

follows: when the number of parts increases, the difficulty of (randomly) producing a blend with adequate fitness also increases. As such, blends with fewer exchanges of parts, thus closer to base representation (in which all the relations are satisfied), tend to become dominant during the initial generations of the evolutionary runs. We consider that a significantly higher number of runs would be necessary to produce blends with more exchanges. Furthermore, valuing the exchange of parts, through the modification of the fitness function, may also be advisable for promoting the emergence of such blends.

Nevertheless, it is also important to consider that this feature is in agreement with the *Topology principle* – from the *Optimality Principles* (see Section 4.1.2) presented by Fauconnier and Turner (1998) – which values the similarity degree to the input spaces (in this case, with one of the base representations). As the blends are produced as visual representations that work as wholes as well as sets of individual parts, the *Principle of Integration* is respected by design – from the *Optimality Principles* presented by Fauconnier and Turner (1998).

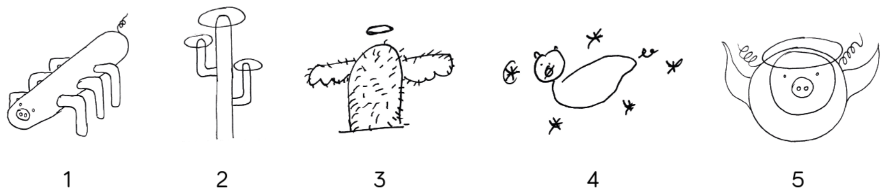


Figure 6.11: Hand-drawn blends – 1 corresponds to *pig-cactus*, 2-3 to *angel-cactus* and 4-5 to *pig-angel*.

6.3.1 *Comparison with user-drawn blends*

As described in Section 6.2.1.3, during the initial phase of the project we conducted a task of collecting visual blends drawn by the participants. A total of 39 drawn blends were collected, from which 14 correspond to the blend between *cactus* and *angel*, 12 correspond to the blend between *cactus* and *pig* and 13 correspond to the blend between *pig* and *cactus*.

The implemented blender was able to produce visual blends similar to the ones drawn by the participants (Fig. 6.10). After analysing the produced blends, the following results were obtained:

- 23 from the 39 drawn blends (DB) were produced by our *Blender*;
- 2 are not possible to be produced due to inconsistencies (e.g. one drawn blend from *angel-pig* used a mapping from *wing-tail* and at the same time maintained the *wings*, 5 in Fig. 6.11);
- 6 were not able to be produced in the current version due to mappings that were not produced by the *Mapper* (e.g. *head* from *angel* with *body* from *cactus*, 2 and 3 in Fig. 6.11);
- 5 were not able to be produced because not all of the collected drawn representations were used in the experiments (no star elements were used, which are needed to produce 4 in Fig. 6.11).

According to the aforementioned results, the implemented *Blender* is not only able to produce blends that are coherent with the ones drawn by participants but is also able to produce novel blends that no participant drew, showing creative behaviour.

6.4 EVALUATING PERCEPTION OF PRODUCED BLENDS

In order to assess if the produced blends could be correctly perceived, a second enquiry was conducted. The main goal was to evaluate whether the participant could identify the input spaces used for each blend (i.e. if it was possible to identify *pig* and *cactus* in a blend produced for *pig-cactus*). This is related to the *Unpacking Principle* (Fauconnier and Turner, 1998) (see Section 4.1.2).

Figure 6.12: Examples of produced blends.

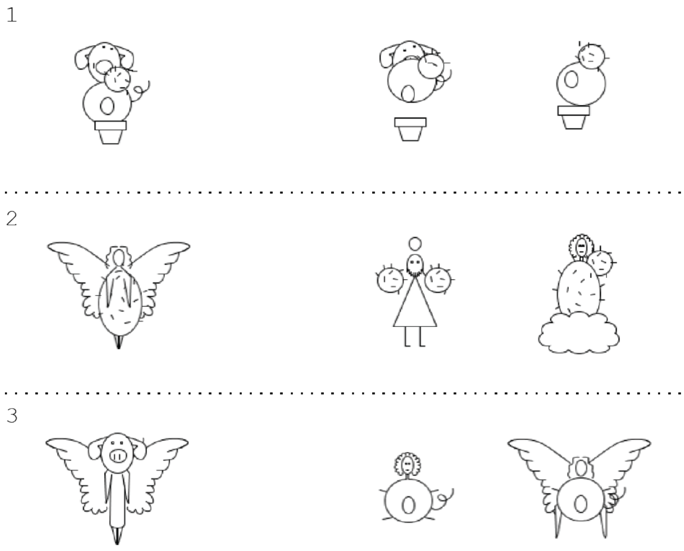


Figure 6.13: Examples of the visual blends presented in the second enquiry. On the left are the "good" blends (one for each) and on the right are the "bad" blends (1 corresponds to *cactus-pig*, 2 to *angel-cactus* and 3 to *angel-pig*).

6.4.1 *Experiment Setup*

In the first enquiry (described in Section 6.2.1.3), the fourth task (T4) consisted in collecting the *prototypical parts* for each concept – these are the parts that most identify the concept (e.g. *wing* for *angel*). We used the data collected to produce the second enquiry. First, we evaluated the quality of produced blends based on two criteria: *fitness* of the individual and *presence* and *legibility* of the *prototypical parts* (i.e. a "good" exemplar is an individual with the prototypical parts clearly visible; a "bad" exemplar is an individual with reduced number or absence of prototypical parts). For each blend (*angel-pig*, *cactus-pig* or *angel-cactus*), four visual blends were selected (two considered "good" and two considered "bad", see Fig. 6.13).

A total of 12 visual blends were used (six "bad" and six "good"). In order to minimise the biasing of the results, each participant evaluated two visual representations (one "bad" and one "good") of different blends (e.g. when the first was of *cactus-pig*, the second could only be of *angel-pig* or *angel-cactus*). The "bad" blends were evaluated first to further reduce bias.

6.4.2 *Results*

The enquiry was conducted with 30 participants, which resulted in each visual blend being tested by 5 participants. The results (Table 6.2 and Table 6.3) clearly show that the "good" blends were easier to

Table 6.2: Results of the number of correct input spaces' names given for each of the concept combinations, divided by good and bad blends and expressed in percentage of answers.

		0	1	2
cactus-pig	Good	20%	50%	30%
	Bad	50%	50%	0%
angel-pig	Good	10%	20%	70%
	Bad	40%	50%	10%
angel-cactus	Good	0%	60%	40%
	Bad	10%	80%	10%

Table 6.3: Results of the number of correct input spaces' names given for each of the blends used in the experiment (#B), expressed in number of answers.

		#B	0	1	2
cactus-pig	Good	1	1	2	2
		2	1	3	1
	Bad	3	1	4	0
		4	4	1	0
angel-pig	Good	5	0	1	4
		6	1	1	3
	Bad	7	2	3	0
		8	2	2	1
angel-cactus	Good	9	0	4	1
		10	0	2	3
	Bad	11	0	5	0
		12	1	3	1

be correctly named (the percentage of total correct naming is always higher for the "good" examples and the percentage of total incorrect naming is always higher for the "bad" blends). In addition to this, the names of the input spaces were also easier to be identified in some of the representations than in others (e.g. the "good" blends for *angel-pig* received more fully correct answers than the rest of the blends, as shown in Table 6.3).

Overall, the majority of the participants could identify at least one of the input spaces for the "good" exemplars of visual blends. Even though some of the participants could not correctly name both of the input spaces, the answers given were somehow related to the correct ones (e.g. the names given for the input spaces in the first "bad" blend of 3 in Fig. 6.13 were often *pig* and *lady/woman*, instead of *pig* and *angel* – this is likely due to the fact that no *halo* nor *wings* are presented).

6.5 SUMMARY

In this chapter, we presented a descriptive approach for the automatic generation of visual blends. The approach uses *structured visual representations* along with *sets of visual relations* which describe how the elements – in which the visual representation can be decomposed – relate to each other, for example *element A* inside *element B*. These relations are the base for the construction of the visual representations.

We described the implementation of a system (the *Blender*) that is composed of two components: the *Mapper* and the *Visual Blender*. The system can be considered a *hybrid blender*, as the blending process starts at the *Mapper* (*conceptual level*) and ends at the *Visual Blender* (*visual level*). We use an evolutionary engine based on a *genetic algorithm*, and we employ a multi-population setting, with each population corresponding to a different analogy and each individual being a visual blend. The evolution is guided by a fitness function that assesses the quality of each blend based on the satisfied relations.

The results show the ability of the *Blender* to produce *analogies* from *input mental spaces* and generate a wide variety of *visual blends* based on them. The *Visual Blender* component, in addition to fulfilling its purpose, is able to produce interesting and unexpected blends. We compared the blends produced by the system with blends drawn by users to assess the performance of the system. Our comparison demonstrated that the system is able to generate blends similar to the ones produced by users and novel ones. We also assessed the quality of the produced blends by conducting a user study focused on perception, analysing how users name the blends produced by the system, which showed that blends are easier to be perceived when they depict *prototypical parts*.

Based on the conducted experimentation, several possible future enhancements were identified:

(i) exploring an island approach in which exchange of individuals from different analogies may occur if they respect the analogy of the destination population;

(ii) exploring the role of the user (*guided evolution*), by allowing the selection of individuals to evolve;

(iii) considering *Optimality Principles* in the assessment of fitness (e.g. how many parts are exchanged) and exploring which of them may be useful or needed – as discussed by Martins et al. (2016);

(iv) using relations such as *biggerThan* or *smallerThan* to explore style changing (e.g. the style of the produced blends will be affected if a base visual representation has *head* biggerThan *body*);

(v) exploring context in the production of blends (e.g. stars surrounding the angel).

The system described in this chapter produces visual representations for the combination of concepts that are given as input, through a hybrid process of blending (*conceptual* and *visual*). One of the main goals of this exploration was to assess how structured visual representations could be used to facilitate the blending process, which in turn could be used for visual representation of concepts. In terms of this thesis' goals, the capability of producing visual representations for concepts is limited by the input concepts, only producing representations for their combinations. Moreover, the system relies on specific input, requiring *input spaces* (in the form of *semantic networks*) and *visual representations* (in SVG). In addition, for each input concept, these two types of input need to be aligned for the blending process to work. In this sense, the system is limited in terms of *conceptual reach* and dependent on prior production of materials for the input concepts.

Part III

EMOJI

In this part of the thesis, we present our research on the main case study: *Emojinating*.

First, we introduce *emoji* and existing research related to them. Second, we propose that emoji are especially suitable to be used in visual blending. Then, we describe the implementation and evaluation of the different versions of *Emojinating* – a system that uses visual blending of emoji and semantic network exploration to generate visual representations for concepts introduced by a user.

We use the *Emojinating* system as a case study to analyse the appropriateness of visual blending for the visual representation of concepts. We conduct several experiments in which we analyse *output quality*, *type of blend* used, *usefulness* to the user and *easiness of interpretation*.

The results produced by *Emojinating* vary in terms of *conceptual complexity* (representing abstract and concrete concepts) and *nature of representation* (going from literal to non-literal). We believe that the system has the potential to be explored as an ideation-aiding tool to be used in brainstorming activities, by presenting the user with visual representations of the concepts given as input.

7 EMOJI: PICTURE-WORDS TO PICTURE CONCEPTS

This chapter introduces *emoji*, describes existing research topics and presents the characteristics of emoji set that make it a suitable asset for *visual conceptual blending*.

7.1 EMOJI AS PICTURE-WORDS

The word *emoji*[1] has a Japanese origin, in which the *e* (絵) means "picture" and *moji* (文字) means "written character"[2] – leading to a possible interpretation of emoji as "picture-word".

The first set of emoji consisted of 176 images and was designed by Shigetaka Kurita in 1999 for the Japanese telecommunication organisation *NTT DoCoMo*[3] (Fig. 7.2). In 2010, emoji characters were codified by the *Unicode Consortium*[4] – a non-profit organisation that develops and maintains the text encoding standard known as the *Unicode Standard*, which includes every character used by modern software and information technology protocols (Lucas, 2016). Once codified, they became usable in different platforms and languages (see Fig. 7.1). In 2011, *Apple* included an emoji keyboard in iOS5 and *Android* followed in 2013. These developments contributed to the worldwide adoption of emoji, which became more popular than their predecessor *emoticon* – sequences of ASCII characters often used to express emotions in Computer-Mediated Communication (CMC) (e.g. ":)").

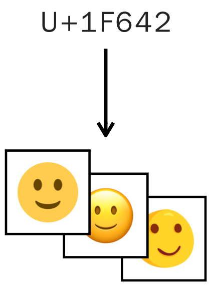


Figure 7.1: Codepoint assigned by *Unicode* to the *Slightly Smiling Face* emoji and its versions in different platforms.

Nowadays, emoji are an important part of our way of writing. Their increasing relevance is easy to observe: they have received attention from language-related resources (*Oxford Dictionaries* named the emoji *face with tears of joy* 😂 (`U+1F602`) the Word of The Year 2015[5]); their usage is on a rising trend – *Facebook* reported in 2017 that 60 million emoji are used every day on *Facebook* and five billion on Messenger[6] and *Emojipedia* reported that in July 2021 more than one in five tweets contain an emoji;[7] and new emoji-related tools or features are constantly being developed (e.g. new *Memoji* customisation options introduced in 2020 with iOS14,[8] such as new hairstyles, headwear, or masks). According to a report released by Adobe (2019), users surveyed admit to

1 the term "emoji" is used for both singular and plural
2 unicode.org/reports/tr51/proposed.html
3 edition.cnn.com/style/article/emoji-shigetaka-kurita-standards-manual/
4 unicode.org/
5 en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2015
6 blog.emojipedia.org/5-billion-emojis-sent-daily-on-messenger/
7 blog.emojipedia.org/emoji-use-at-all-time-high/
8 www.macrumors.com/how-to/create-memoji-ios-12/

Figure 7.2: *NTT DoCoMo's* original set of emoji designed by Shigetaka Kurita, currently part of MoMA's collection.

include emoji in text messages 49% of the time. Despite being mostly used as a way to make conversations more fun or lighten the mood, 94% of the surveyed users identify the ability to communicate across language barriers as one of the greatest benefits of emoji.

As of *Unicode Emoji 13.1*, approved in September 2020, there are more than 3500 different emoji. They vary in nature as the emoji set includes *pictograms* (e.g. *sun* `U+2600` ☀️), *ideograms* (e.g. *no littering* `U+1F6AF` 🚯) and *logograms* (e.g. *japanese "reserved" button* `U+1F22F` 🈯). They also vary in terms of *concreteness*, going from *concrete* (e.g. *lion* `U+1F981` 🦁) to *abstract* (e.g. *curly loop* `U+27B0` ➰).

The high *conceptual reach* of emoji allows them to be used in different ways and for different purposes in written communication. Some authors even discuss a shift towards a more *visual language* (Danesi, 2017; Lebduska, 2014). This shift would in fact bring us close to old ways of writing, such as *hieroglyphs*.[9] The integration of emoji in written language is easy to observe by considering existing features that connect emoji and text. Some examples are *Search-by-emoji* supported by *Google*,[10] and the *Emoji Replacement* and *Emoji Prediction* features available in iOS 10.[11] Danesi (2017) refers to the combination of text and emoji as "hybrid emoji writing", seen as an amplification of traditional writing in which emoji often add, clarify and reinforce meaning. This way, the connection between emoji and their assigned meanings makes them much more valuable than simple decorative characters. In particular, emoji are often used to express emotions, which makes them relevant to understand the meaning of written messages.

9 The hieroglyphic writing is not purely pictorial, instead it is a mixture of both phonographic and ideographic components (Jespersen and Reintges, 2008).

10 `forbes.com/sites/jaysondemers/2017/06/01/could-emoji-searches-and-emoji-seo-become-a-trend/`

11 `macrumors.com/how-to/ios-10-messages-emoji/`

7.2 EMOJI AS A RESEARCH FIELD

Ever since they were introduced to the world, emoji have been attracting people's attention, not only from the general public, who frequently uses them in written text but also from researchers who consider that emoji are worth studying. We identify six main topics addressed in emoji-related research: *Meaning*, *Sentiment*, *Interpretation*, *Role in communication*, *Similarity* and *Generation*.

Research on emoji meaning mostly focuses on the application of word embedding techniques to data from different sources (e.g. Barbieri, Ronzano, and Saggion, 2016; Dimson, 2015; Eisner et al., 2016). The sentiment of emoji is often inferred based on the sentiment of the surrounding text (e.g. Novak et al., 2015) and has been used to study emoji usage intentions (Hu et al., 2017). Miller et al. (2016) study how emoji rendering variation between platforms affects their interpretation, and Rodrigues et al. (2018) delve into the differences between user interpretation and the meaning intended by the developer.

Concerning the role of emoji in written communication, several topics have been addressed: redundancy and part-of-speech category (Donato and Paggio, 2017), complementary vs text-replacing functions of emoji (Dürscheid and Siever, 2017), emoji as text-replacement and its effect on reading time (Gustafsson, 2017), emoji as semantic primes (Wicke, 2017), among others (Cramer, Juan, and Tetreault, 2016; Herring and Dainas, 2017; Kelly and Watts, 2015).

Emoji similarity can be interpreted as either visual or conceptual. Despite this, most research focuses on the latter. Ai et al. (2017) semantically measure emoji similarity. Some authors employ vector embeddings to identify clusters of similarity (Barbieri, Ronzano, and Saggion, 2016; Eisner et al., 2016). Pohl, Domin, and Rohs (2017) apply agglomerative clustering to organise emoji into a relatedness-hierarchy. Wijeratne et al. (2017a) introduce *EmoSim508*, a dataset of human-annotated semantic similarity scores.

On emoji generation, existing research mostly focuses on the use of Generative Adversarial Networks (GANs). Radpour and Bheda (2017) propose an approach to generate emoji from text that uses a deep convolutional GAN conditioned on word vectors from *Google*'s *word2vec* model and trained with face emoji and words assigned to them. Puyat (2017) follows a similar approach by training a conditional GAN on 2,000 emoji images and names scraped from `unicode.org` and conditioned on word vectors. Mittal et al. (2020) train a GAN based model on sketch-emoji pairs, which allows them to produce new emoji-like images from crude sketches. They also experiment with multimodal input, enabling the system to generate an image from a partial sketch and a handwritten keyword (e.g. "happy"). All these approaches focus on face emoji (e.g. "smiley face") and result in outputs whose quality is significantly lower when compared to official emoji. Radpour and

Bheda (2017) position their future work in alignment with ours by mentioning that a future use of their *text-to-emoji* approach is to produce representations of novel concepts (e.g. *food coma*) through the combination of existing emoji, focusing on certain features.

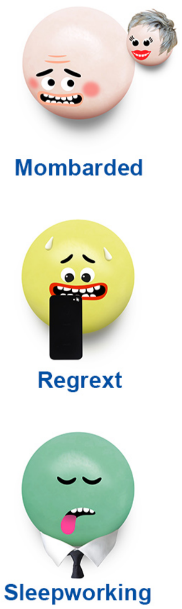


Figure 7.3: *Mentos' Ementicons* created in 2015.

Figure 7.4: "Bootlicker emoji" from (Sittenfeld and Daniel, 2014).

7.3 An increasing lexicon

The constant desire to represent new things is one of the main motivations behind emoji. Each year new emoji are added to *Unicode*, increasing its conceptual reach. These new additions are the result of a selection process in which emoji proposals are analysed by the *Unicode Emoji Subcommittee* and selected based on a set of *Selection Factors*.[12]

7.3.1 *Proposing new emoji*

Despite this constant addition of new emoji, there are still a large number of concepts that are not represented. Several attempts have been made to complement the emoji lexicon. The nature and goals of such attempts are not always the same. Some examples are: to propose culture-specific emoji, e.g. Finland emoji;[13] to include a certain trait, e.g. curly hair (Neff, 2015); to help abuse victims;[14] or even to propose "missing emoji", e.g. *menstruation* (Ho, 2019).

Although the submission of emoji proposals is open to the general public, only a small percentage are accepted for encoding. The selection factors are most welcoming of emoji that are not too specific and may have multiple usages (e.g. metaphorical references or symbolism), and that do not overlap with existing emoji.

The absence of more abstract concepts in the emoji set is especially evident. Despite this, there is interest in the representation of less concrete and also more creative concepts. This can be observed in advertising campaigns that take advantage of this gap in the emoji set, e.g. the *Ementicons* by *Mentos*,[15] which proposed several emoji-like icons, e.g. *sleepworking* (see Fig. 7.3). Another example is a list of emoji proposed in the article *The Emojis We Really Need* published in *The New York Times* (Sittenfeld and Daniel, 2014), e.g. the emoji to use when you like someone's tweet because they are in a position to help you professionally (see Fig. 7.4). A more serious list is the one proposed by *Microsoft* (see Fig. 7.5), some of which were eventually added to the emoji set (e.g. *t-rex* `U+1F996` 🦖 or *lobster* `U+1F99E` 🦞). This serves to show that the general public also values the visual representation of unusual and culture-related concepts.

12 unicode.org/emoji/proposals.html
13 finland.fi/emoji/
14 webcollection.se/bris/abusedemojis/
15 www.fastcompany.com/3043786/now-you-can-use-mentos-emoticons-to-show-how-you-really-feel

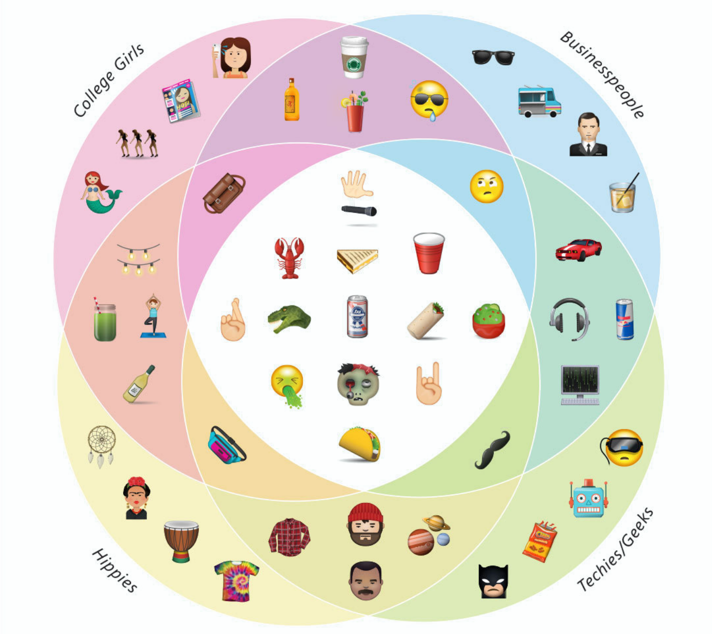


Figure 7.5: 49 emoji that should exist according to *Microsoft*.

7.3.2 *Semantic change in Emoji*

One interesting side of emoji is that their meaning is subject to change. This change is usually caused by the need to have a representation for a concept that does not have an emoji. For example, the absence of sex-related emoji led to the creative use of the *eggplant* (`U+1F346`) to represent male genitalia and the *peach* (`U+1F351`) to represent buttocks. The repurposing of emoji is studied by Wiseman and Gould (2018).

A recent example is the *syringe* emoji (`U+1F489`), which was initially implemented for blood donation. Based on an analysis of tweets from 2018 and 2019 by *Emojipedia*,[16] it was used together with words such as "blood", "veins", "inject" and also "tattoo". With the COVID-19 outbreak and the hope for a vaccine, the *syringe* emoji started to be used to represent a different concept: *vaccination*. In the same study, *Twitter* data from December 2020 and January 2021 shows that tweets using the *syringe* emoji are more focused on topics related to COVID-19 vaccines (e.g. "vaccine", "COVID" or "Pfizer").

A study by Robertson et al. (2021) focuses on the topic of semantic change, which is defined as the development of meaning over time. Robertson et al. (2021) employ semantic change detection techniques to emoji and show that emoji can undergo semantic change over time.

Also interesting is how the standard design can suffer changes to allow for more versatility. In the case of *syringe* emoji, we can see how the emoji vendors were converging into a similar design (2015-2020 in Fig. 7.6), which had blood inside the syringe and sometimes blood

16 blog.emojipedia.org/vaccine-emoji-comes-to-life/

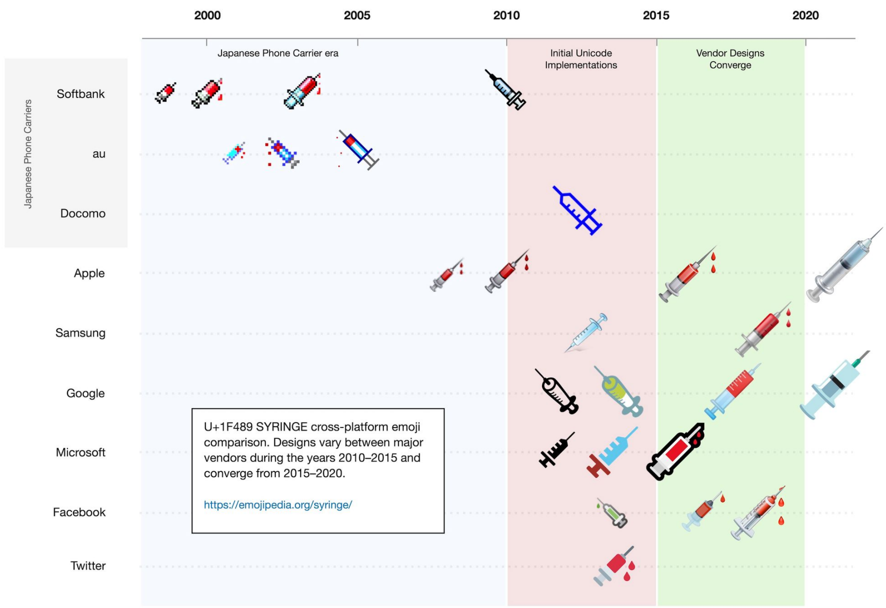


Figure 7.6: Cross-platform comparison of *syringe* emoji (`U+1F489`) 1999-2021. Source: *Emojipedia*.

drops. However, for uses related to vaccination, showing blood in the syringe was a limitation. As of 2021, we see *Apple*'s and *Google*'s designs change into a liquid-free and vaccine-friendly syringe emoji 💉.

7.4 CUSTOMISATION

In addition to the constant wish to cover more and more concepts, there is another craving from users in regards to emoji: *customisation*. This has long been a functionality explored to increase users' interest. In the early 2000s, *Windows Live Messenger*[17] allowed the user to create emoticons by uploading an image file. *Slack*[18] currently has the same feature. Other applications allow face-related customisation, e.g. *Bitmoji*.[19]

Researchers have also devoted some attention to customisation. One example which is related to variation is presented by Barbieri et al. (2017), who investigate the properties of derivations of the *kappa emote* in *Twitch*. Taigman, Polyak, and Wolf (2016) transform photos of faces into cartoons and, as previously mentioned, Mittal et al. (2020) produce emoji-like images from crude sketches, having customisation as the goal.

As the customisation of emoji characters is not possible, the majority of customisation examples concern other emoji-like figures. An exam-

17 news.microsoft.com/2003/06/18/msn-messenger-6-allows-im-lovers-to-express-themselves-with-style/
18 get.slack.help/hc/en-us/articles/206870177-Create-custom-emoji
19 bitmoji.com

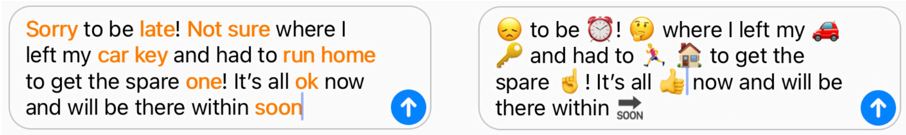


Figure 7.7: *Emoji Replacement* feature in *Apple* iOS.

ple is the platform *emoji.gg*, which has a feature called *maker*[20] that allows users to combine emoji parts (e.g. bases, eyes, brows, mouths, etc.) and produce stickers. Moreover, *Apple* addressed the customisation issue by presenting users with *Animoji* (released in 2017) and *Memoji* (released in 2018),[21] which follow the design of *Apple*'s emoji but have customisation options and can be used to mirror one's facial expression through facial recognition. They can be sent as stickers or video. The fact that they are not actual emoji does not seem to matter to users, as long as they are able to somehow send them.

In 2018, *Google* launched *Emoji Minis*, which allow users to produce emoji-like versions of themselves.[22] In early 2020, *Google* introduced *Emoji Kitchen*,[23] a *Gboard* feature that gives users the freedom to combine emoji. The feature relies on previously drawn stickers and presents them to the user as *mashups*. In Chapter 5, we described an analysis of *Emoji Kitchen* blends.

It is important to mention that even though the results of these examples are visually similar to emoji, they are not encoded by *Unicode* and, for that reason, cannot be called emoji. Until a proposal is accepted by *Unicode*, the "emoji" cannot be uniformly used by different platforms.

7.5 Emoji for Concept Representation

In regards to using images to convey concepts or ideas, emoji are a key example. Due to their wide conceptual reach, emoji can be used for different purposes. Two different functions of emoji are identified by Dürscheid and Siever (2017): *complementary*, by accompanying texts usually added at the end of sentences; and *replacement*, by substituting them into sentences. The use of emoji as complementary signs in written communication enriches it (Niediek, 2016) by allowing the transmission of *non-verbal cues*, e.g. *face expressions*, *tones* and *gestures* (Hu et al., 2017), which are lacking in written communication and CMC.

On the other hand, emoji are often used as replacements. Several features foster this type of function by making it easy for the user to make substitutions (e.g. the emoji suggestion in iOS, Fig. 7.7). This way of

20 emoji.gg/maker
21 blog.emojipedia.org/ios-13-adds-memoji-to-emoji-keyboard/
22 techcrunch.com/2018/10/30/googles-gboard-now-lets-you-create-a-set-of-emoji-that-look-like-you/
23 blog.google/products/android/feeling-all-the-feels-theres-an-emoji-sticker-for-that/

using emoji brings other aspects into play, such as *reading time* and *message comprehension*. Existing research provides evidence that reading time increases but comprehensibility does not seem to be negatively affected (Cohn et al., 2018; Gustafsson, 2017). These results apply to multimodal integration and not to full replacement. An example of the latter is the translation of *Moby-Dick* by Herman Melville into emoji – *Emoji Dick*,[24] produced using human crowd-workers.

An ongoing discussion concerns whether emoji can constitute a language. The idea of emoji as language has been mentioned by several authors (e.g. Danesi, 2017; Lebduska, 2014) but there are different views on the topic. For example, Cohn, Engelen, and Schilperoord (2019) state that even though emoji are an effective communicative tool, they lack when it comes to grammar by having limited complexity. In contrast, Ge-Stadnyk (2021) and Herring and Ge (2020) provide evidence that emoji usage is increasingly taking on characteristics of verbal language, e.g. the subject being syntacticised as *clause-final* position, and mention the possibility of emoji being an "emergent graphical language".

In any case, there is no doubt that emoji can be used in multiple ways to represent concepts. First, most emoji have an iconic relationship with the referents that they designate, e.g. 🍺 for *beer mug*. Second, they can be considered to represent a class of entities much larger than a specific image, e.g. 🍺 not only represents a mug with the emoji's exact shape, filled with beer of exactly the same colour, but it also represents beer in general. Another example is the emoji for *aeroplane* (`U+2708`) ✈️, which can refer to the plane itself, to flying or travelling and even an airport. As such, emoji can be understood as semiotic objects (Danesi, 2017), given that they can be interpreted literally, metaphorically and even symbolically. This versatility is intentional, as *Unicode* considers emoji as building blocks.[25] A similar view is proposed by Wicke (2017) by studying emoji as semantic primes. As such, the potential for concept representation goes beyond single characters.

Wicke (2017) proposes a system to translate action words into sequences of emoji through the use of linguistic strategies (e.g. *metaphor*, *idioms*, *rebus*, etc.). Through an empirical evaluation, Wicke (2017) concludes that the strategies that lead to the best understood and appreciated translations are the *rebus principle* (e.g. 🐝🍁 for *believe*), *metaphors* (e.g. ☘️ for *luck*) or *literal translations* (e.g. 💥 for the action *to explode*). According to Wicke (2017), these strategies enable one to use the semiotic advantages of emoji.

Wicke and Bolognesi (2020) conduct a user study in which they asked participants to provide semantic representations for a sample of 300 English nouns using emoji, with the goal of identifying which representational strategies are most used to represent abstract and concrete concepts. They use a refined version of the classification of repre-

24 `www.emojidick.com`
25 `unicode.org/emoji/proposals.html`

sentational strategies proposed by Wicke (2017): *literal*, *rebus*, *phonetic similarity* and *figurative construction*. According to their results, *figurative construction* is the most used strategy (59%), followed by *literal* (33.91%). They conclude that when there is not an emoji that represents the concept literally, users resort to figurative constructions. They also conclude that abstract concepts require a higher number of emoji to be represented, as people tend to represent situations in which concept is experienced. This contrasts with concrete ones in which people focus on the entity designated by the concept.

Examples of the representation strategies are:[26]

- *Literal*: the emoji represents the actual referent (e.g. 🦁 for *lion*);
- *Rebus*: emoji denote words that compose the sound of the target word when read aloud (e.g. 🐝🍁 for *believe*);
- *Phonetic similarity*: emoji denotes a word that is phonetically similar to the target word (e.g. 🐚 for *shelf*);
- *Figurative Construction*: the emoji has a connection to the target word based on symbols (e.g. ☮️ for *peace*), metaphors (e.g. ☘️ for *luck*) and metonymies (e.g. 🔒 for *cage*).

In addition, the representation of some concepts may require the use of several emoji, which have to be all considered to construct the meaning – e.g. *garage* requires the combination of a *car* and a *house*. Some types of emoji sequencing are:

- *Plural*: 🍺🍺 for *beers*;
- *Category*: 🍸🍹🍺 for *drink*;
- *Time or cause-effect relation*: 📘🛂🛫 for *passport*;
- *Spatial relation*: 🚗🏠 for *garage* or 🌊🌴 for *beach*.

Similarly, Cohn, Engelen, and Schilperoord (2019) provide an extensive list of different types of emoji sequencing, some of which were already mentioned (e.g. *rebus* or *temporal sequence*). Two of them are especially worth mentioning as they are more related to the creation of a combined image rather than a sequence. The first is *affixation*, in which two emoji are put together to produce a larger unit, e.g. in 🏃💨 the puff of smoke emoji is used as a *visual affix* to enhance the sense of speed. The second, named *whole image*, consists in combining emoji to produce a single *picture*, e.g. ⛰️🏃 represent a person running towards a mountain. These last two strategies come closer to what can be considered *visual blending*.

26 These representation strategies are related to the *Types of Correspondence* of signs addressed in Section 2.3.2.2.

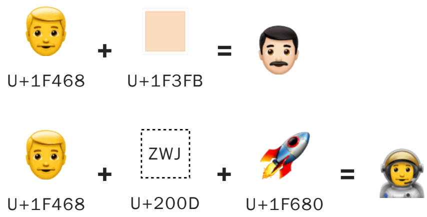


Figure 7.8: *Modifier* (top) and *Zero-Width-Joiner* mechanisms (bottom).

7.6 VISUAL BLENDING OF EMOJI

In addition to being used in sequences to represent meaning, emoji may also be used to produce visual representations for new concepts. This can be observed in some of the examples previously given, such as the *sleepworking Ementicon* (see Fig. 7.3) in which a person sleeping is put together with a shirt and a tie to represent sleeping on work. An interesting fact is that this connection to visual blending is not new.

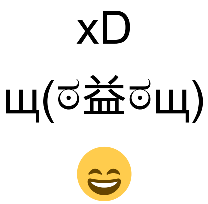


Figure 7.9: "Laughing" western emoticon (top), "Why" eastern emoticon (middle) and "Grinning Face With Smiling Eyes" emoji (bottom).

7.6.1 *Connections to Visual Blending*

Before emoji, sequences of ASCII characters were often used to express emotions in CMC – they are called *emoticons* (see Fig. 7.9). One of the emoticons' advantages is their potential for *customisation* and *variation*. Whereas emoji are inserted as a whole in the text, emoticons are the result of a combination of individual components (Dürscheid and Siever, 2017) – e.g. ":" + ")" = ":)". The changeable parts not only allow a high degree of visual variability but also the exchange of a component leads to a change in the meaning. This is one of the reasons why they are still being used as alternative to emoji (Guibon, Ochs, and Bellot, 2016).

In 2015, the *Unicode Consortium* decided to add *skin tone modifiers* (characters that modify other emoji) to *Unicode* core specifications (see Fig. 7.8). One year later, the Zero-Width-Joiner (ZWJ) mechanism was also implemented – an invisible character to denote the combination between two characters (Abbing, Pierrot, and Snelting, 2017). This development meant that new emoji could be created by combining others (see Fig. 7.8).

There are also examples of systems that generate emoji blends. To the best of our knowledge, the earliest example is *Emojimoji*.[27] *Emojimoji* is an emoji generator that randomly merges emoji shapes and names (Fig. 7.10). It was implemented as a research project of the group *Emblemmatic*. *Emblemmatic* is presented as having the goal to use computational techniques to explore symbols and their meanings, which is fully aligned with the topic of this thesis. Despite this, we consider that the conceptual side of *Emojimoji* is reduced as it only uses *semantic knowledge* for name combination.

Figure 7.10: *Emojimoji* interface, showing a combination of two emoji and their names.

27 emblemmatic.org/emojimoji

In 2019, a *Twitter* bot called *Emoji Mashup Bot* (*@EmojiMashupBot*) was released. The bot produces combinations of two emoji, using previously selected pieces – e.g. when combining two smileys, it takes the eyes and mouth of one and the base of the other. A few months later, another *Twitter* bot was introduced by the same author: the *Emoji Mashup Bot+* (*@EmojiMashupPlus*). This second bot works differently, it uses three lists of emoji parts (*base*, *eyes* and *mouth*) and randomly selects one element from each list to produce a blend. There is also a webpage[28] where users can produce their own combinations. The author of the bots also released a sticker pack[29] of 29 blends. This sticker pack was in a way the predecessor of the already mentioned *Gboard Emoji Kitchen* feature, which also has a webpage[30] for emoji combination. While the stickers from the *Emoji Mashup Bot* are automatically generated, the ones from *Emoji Kitchen* are designed by hand.[31]

All in all, none of the mentioned approaches uses semantic knowledge in emoji generation, which is one of the main goals of our work.

7.6.2 *The output of emoji blending is not an emoji*

Despite all these examples of visual blending of emoji, one thing needs to be made clear: the actual combinatorial possibilities of emoji are limited by their encoding as single characters, as pointed out by Cohn, Engelen, and Schilperoord (2019). In contrast with emoticons, users do not have control over the several parts of emoji. As such, users cannot, currently, create or change existing emoji.

Moreover, by using emoji as characters, they are constrained to be placed like text in a linear sequence. This does not allow a combination in a natural way. An example given by Cohn, Engelen, and Schilperoord (2019) is *dumpster fire*, which, if freely drawn, would most likely result in a fire flame on top of the dustbin. However, by using emoji according to their expected way of use, these can only be placed in a linear sequence: 🗑️🔥.

To address this issue, one can choose to approach emoji from a different angle. Being encoded by *Unicode* and used as whole units, emoji are considered characters. On the other hand, they are designed as pictures rather than typographic signs. In fact, each emoji has different versions, as each vendor is responsible for how it is rendered on its platform. By taking this perspective, emoji can be studied as pictures and not as characters, opening other possibilities, such as visual blending. As such, emoji are no longer the goal but only the means to produce something else. This way, any venture into emoji visual blending

28 louan.me/EmojiMashupBot/
29 theverge.com/tldr/2019/8/20/20825652
30 emoji.kitchen
31 www.theverge.com/2020/2/12/21132971

needs to be built on the following premise: the result of a process of visual blending of emoji is not an *emoji*.

7.6.3 *Emoji as Visual Blending Resource*

Using emoji for visual blending draws inspiration from their modifier and ZWJ mechanisms (see Fig. 7.8), implemented by *Unicode*. Although *visual blending of emoji* does not produce *emoji*, we believe that it has a high potential for concept representation, which can be used for *ideation* or *visualisation purposes*. This idea is supported by the previously given examples of more creative emoji-like stickers, some of which were produced using a visual blending approach (see "bootlicker" in Fig. 7.4).

As we previously mentioned in Section 4.3, when implementing a visual blending system there are certain requirements that one needs to meet, especially in regards to input materials (i.e. *images* and *semantic knowledge*). Our position is that emoji possess unique characteristics that make them highly suitable to be used in visual blending. As such, there are several reasons why the emoji set can be considered a useful resource for visual blending:

CONCEPTUAL REACH: The emoji set has a large conceptual reach, focusing on core concepts and covering a long list of topics – e.g. animals, plants, fruits, food, activities, places, objects, etc.

HIGH NUMBER: By considering the emoji list as a list of visual representations, it is composed of a high number of elements (more than 3500 in version 13.1, approved in September 2020). Moreover, new emoji are added every year, making it a dataset in constant development.

SIMPLICITY AND DISTINCTIVENESS: Emoji should be recognisable at small sizes, being used at 18×18 pixels on mobile screens. This means that their design needs to be simple, which requires an effort of optimisation in selecting what is included in the emoji and how. Moreover, one of the selection factors is that they represent a distinct and visually iconic entity, which is recognisable by most people familiar with that entity. As such, emoji design usually focuses on *diagnostic features* of the entity, for example, the *pig nose* when representing a *pig*. These characteristics are key in a process of visual blending, avoiding the need to deal with excessive complexity.

SEVERAL DATASETS: The emoji set is not just one dataset of images but multiple. Each emoji has different versions, as each vendor is responsible for how it is rendered on its platform. As such, several emoji datasets exist, some of which can be freely used – e.g.

Noto Emoji,[32] *OpenMoji*[33] or *Twemoji*.[34] One of the key aspects of these different datasets is that the representations for each emoji tend to converge in configuration, i.e. they are similar in terms of content, except for certain cases, as studied by Miller et al. (2016). On the other hand, each dataset has a specific style that is maintained throughout all emoji, achieving high coherence in style and also good design quality. Nonetheless, for generation purposes, multiple emoji datasets can be used in combination.

FORMAT: Several emoji datasets supply images in multiple formats, including Scalable Vector Graphics (SVG). SVG format facilitates the process of blending due to its layered structure, allowing an easy exchange of parts without the need to resort to image processing techniques.

SEMANTICS: Due to their usage in written communication (Adobe, 2019), they have a strong semantic connection, resulting in a combination of *images* and *meanings*, which can be used to produce blends with specific meanings. As we have previously described, emoji meaning is not only iconic-based but also metaphoric and symbolic. In addition, this semantic knowledge is easy to access, for example with the *EmojiNet* dataset (Wijeratne et al., 2017b), a sense inventory built through the aggregation of emoji explanations from multiple sources.

These characteristics make emoji suitable to be exploited in computational approaches for visual representation of concepts. On the one hand, it can lead to systems that produce blends without a conceptual grounding, such as the *Emoji Mashup Bot*. However, a possible goal would be the implementation of approaches that could produce *visual conceptual blends*. This would require a strong integration of the *visual* and *conceptual* levels. Ideally, such a system would be able to produce results like some of the ones shown in Fig. 7.11, which are combinations of emoji blends and the possible context in which they would be used – "the migraine vibe" requires the understanding of what a migraine is and feels like and the ability to produce a visual analogy; "I predict a nap in my near future" requires that the system is able to connect the "zzz" with sleep, the crystal ball with predictions about the future, and also have information that these predictions are expected to be shown inside the ball (a pattern identified in Section 5.3.2.5).

32 github.com/googlefonts/noto-emoji/
33 openmoji.org/about/
34 twemoji.twitter.com

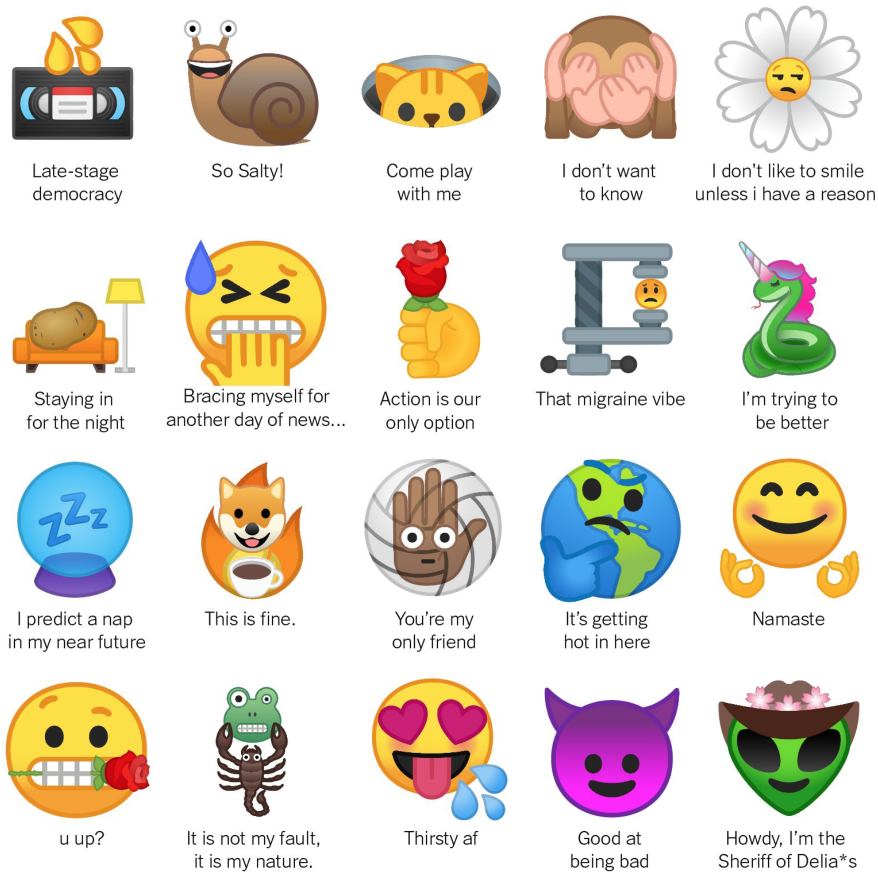


Figure 7.11: *Emoji ligatures* by Jennifer Daniel (`blog.emojipedia.org/hands-on-with-googles-new-emoji-mashups/`)

8 EMOJINATING: GENERATING EMOJI BLENDS

As we have seen in the previous chapter, emoji are well-suited to be used for visual blending purposes. The emoji's connection between *visual representation* and *semantic knowledge*, together with its large *conceptual coverage* have the potential to be exploited in computational approaches to the visual representation of concepts using visual blending. As such, we take advantage of this suitability to implement a computational system that visually represents user-introduced concepts through visual blending – the *Emojinating* system. In this chapter, we describe the first iteration of *Emojinating* and its evaluation.

This chapter is based on the work described in the papers by Cunha, Martins, and Machado (2018a,b).

8.1 IMPLEMENTATION

Emojinating takes inspiration from the combinatorial nature of emoticons and the Zero-Width-Joiner (ZWJ) mechanism (see Fig. 7.8) and makes use of the emoji connection between *pictorial representation* and associated *semantic knowledge* to produce visual representations of concepts through a process of visual blending. The system searches existing emoji semantically related to a user-introduced concept and complements this search with a visual blending process that generates new possibilities by combining emoji.

In this section, we describe the implementation of the first iteration of the system, presenting its *architecture* and explaining the different *system components*.

8.1.1 *Resources used*

In order to build the system, we made use of several emoji-related resources (see Fig. 8.1), which allow us to explore two levels: *visual* and *conceptual*. In the following paragraphs, we describe the resources that were put together when developing *Emojinating*.

8.1.1.1 *Image Dataset*

The first requirement when implementing a visual blending system is the access to a dataset of images. These images are used as input in the visual blending process, in which they are combined to produce new ones. As we have explained in the previous section, there are several image datasets of emoji because each vendor produces its own

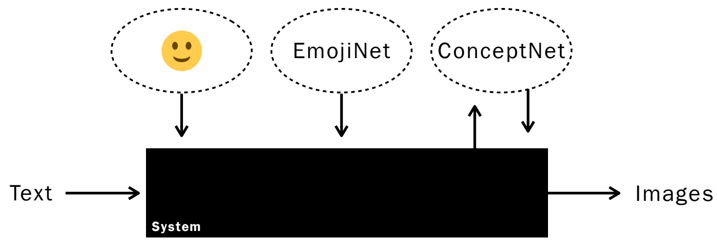


Figure 8.1: *Emojinating* system schematic

version, some of which are free to use. For visual blending purposes, the use of raster images makes it necessary to employ image processing techniques. On the other hand, using an image format in which the different elements are already separated facilitates the process of blending. For this reason, we chose to use datasets that provide images in Scalable Vector Graphics (SVG) format. This decision is aligned with our previous work on the *Pig, Angel and Cactus* experiment described in Chapter 6. The SVG image format enables scaling without reducing quality and uses a layered structure – e.g. each part of an emoji (e.g. a *mouth*) is in a separate layer (see Fig. 8.2). This structure allows an easier blending process and contributes to the overall sense of cohesion among the parts.

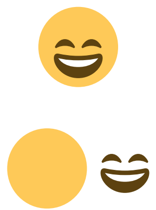


Figure 8.2: Division of "Grinning Face With Smiling Eyes" emoji

Although we experimented with different datasets, most of our work uses *Twitter*'s *Twemoji*,[1] due to style preference. *Twemoji* is a dataset of emoji images made available by *Twitter* (see Fig. 8.3), which provides images in both Portable Network Graphics (PNG) and SVG formats. During the development of *Emojinating*, we used different versions of *Twemoji*, the first was version 2.3, which was available at the time of the development of the first iteration and contained images for 2661 emoji. This dataset consists of images without any semantic information besides the corresponding *Unicode* codepoint in the name of each file.

Figure 8.3: Emoji examples from *Twemoji*

8.1.1.2 *Semantic Knowledge*

Due to their frequent usage in written communication, emoji have semantic knowledge associated with them. This knowledge is useful for tasks such as sense disambiguation, employed to assess the meaning of written text. Having emoji sense disambiguation as a goal, Wijeratne et al. (2016, 2017b) created *EmojiNet*, a machine-readable sense inventory for emoji built through the aggregation of emoji explanations from multiple sources. In *EmojiNet*, each emoji is represented as a nonuple (see Fig. 8.4) that contains data about it (Wijeratne et al., 2017b): *Unicode* representation, name, shortcode, definition, set of keywords that describe attached meanings, set of images from different vendors, set of related emoji, categories that the emoji belongs to and

1 github.com/twitter/twemoji

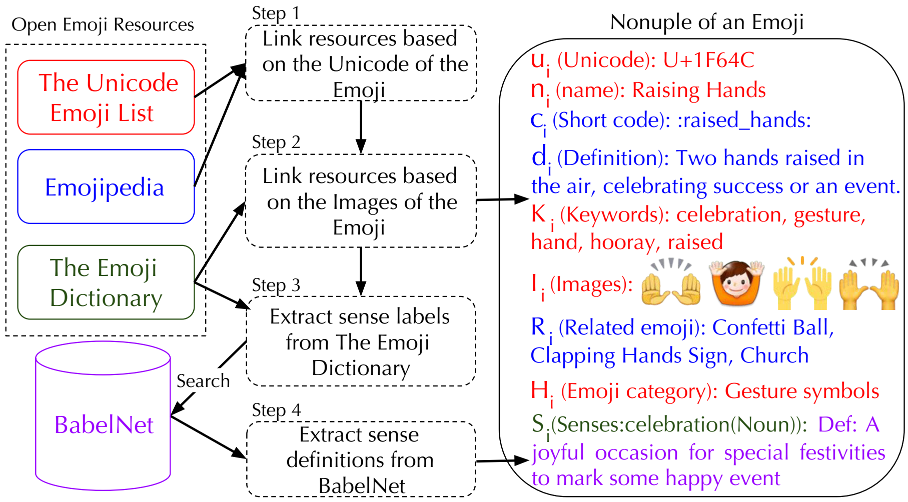


Figure 8.4: Construction of Emoji Representation in *EmojiNet*. Source: (Wijeratne et al., 2017b).

different senses in which the emoji can be used within a sentence. *EmojiNet* contains data of 2389 emoji.

8.1.1.3 *Conceptual Extension*

One of the key aspects to consider when implementing a system for visual representation of concepts is how the system deals with concepts. A particularly important feature is what we refer to as *conceptual extension*: getting information regarding related concepts based on a given concept. To achieve this, we used *ConceptNet* – a *semantic network* originated from the project *Open Mind Common Sense* (Speer and Havasi, 2012). *ConceptNet* provides an Application Programming Interface (API), which can be used to access its data.

8.1.2 *General Architecture*

Even though we focus our attention on the blending of emoji, the main purpose of the *Emojinating* system is to present the user with visual representations for the concept that they introduced. As such, in order to represent the concept, the system firstly searches for *existing emoji* and then produces *visual blends* by combining existing emoji.

For ideation purposes, the blending process is useful if there is no existing emoji that matches the concept but also to suggest possible alternative representations.

The system receives concepts with a maximum length of two words. The system starts by analysing the text given by the user. In this first stage, three things can happen: (i) the user introduces a single word (e.g. *car*), (ii) two words (e.g. *wine polo* or *game theory*) or (iii) more. In the last case, the system removes stop-words (e.g. “a”, “because”,

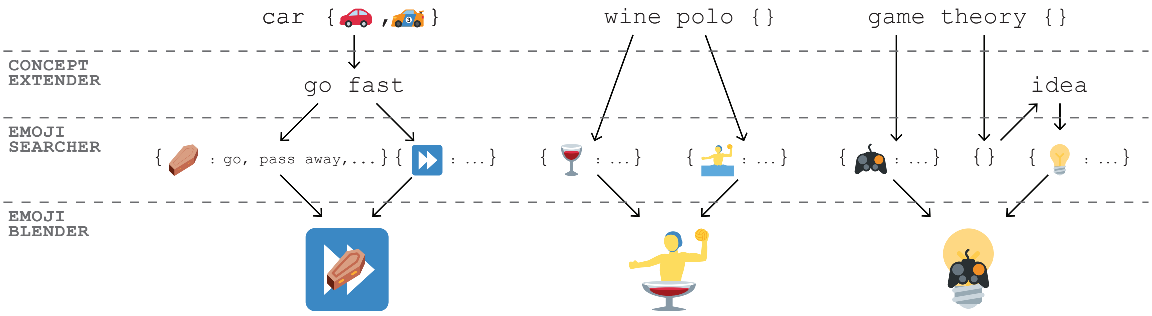


Figure 8.5: *Emojinating*'s process of producing blends for three concepts: *car*, *wine polo* and *game theory*.

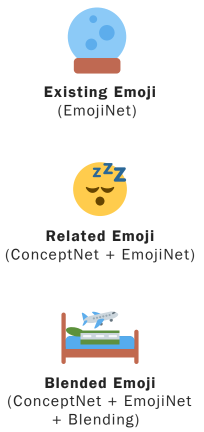


Figure 8.6: Existing emoji, related emoji and blend generated by *Emojinating* for the concept "dream".

"before", "being", etc.) and considers the result as the input text (*query concept*) – if after this process of removal, the word count is still higher than two, the system ignores it and ends the process without any result.

The system conducts two main tasks – retrieval of existing emoji that match the introduced text (T1) and generation of new ones through visual blending (T2). The two tasks are conducted using three components (see Fig. 8.5):

1. *Concept Extender* (CE): uses *ConceptNet* to retrieve concepts related to a given word;
2. *Emoji Searcher* (ES): searches for existing emoji that are semantically related to a given word/words, using semantic knowledge provided by *EmojiNet*;
3. *Emoji Blender* (EB): receives two emoji as input and produces a set of blends.

The output of the system is a set of visual representations for the introduced concept, composed of existing (E) emoji – i.e. emoji that directly match the query word(s) – related (R) emoji – i.e. emoji that match a concept related to the queried one, obtained with *ConceptNet* – and generated blends (B) (see Fig. 8.6).

8.1.2.1 *Retrieval of Existing Emoji* (T1)

In order to conduct T1, the system uses the *Emoji Searcher* component to find emoji based on the word(s) given by the user (e.g. in Fig. 8.5 the *coffin* emoji is retrieved for the word *go* due to its presence in the sense "go, pass away,..."). The word searching process is conducted using the semantic data associated with emoji, provided by *EmojiNet*.

Initially, we conducted a search in senses using their definitions, directly retrieved from *EmojiNet*. However, we concluded that using sense descriptions often leads to unrelated emoji, which are not useful for the system. For this reason, we decided to use the sense lemmas (word(s) that identify the sense, e.g. "celebration" in Fig. 8.4) instead

		
name	woman	woman technologist
keywords	[“woman"]	[“coder", “developer", “software", "woman", "technologist", “inventor"]
definition	“A female face, of adult age. Differentiated from the girl emoji by the lack of pig-tails in her hair. woman …“	"A woman behind a computer screen, working in the field of technology. …“
senses	["woman", "adult_female", "female", …]}	[“woman", "adult_female", "female", “women”, …]
unicode	“U+1F469”	"U+1F469 U+200D U+1F4BB”
	total sense number 7	total sense number 5

Figure 8.7: Two emoji retrieved by using the word “woman”

of their descriptions (e.g. “A joyful occasion...” in Fig. 8.4). As the *EmojiNet* dataset only includes the sense id and its descriptions, the lemmas for each sense id had to be gathered from *BabelNet* (Navigli and Ponzetto, 2012), which was the original source of the *EmojiNet* sense data (Wijeratne et al., 2017b).

The search is conducted in the following data: *name*, *definition*, *keywords* and *sense lemmas*. For the search, an exact matching of the word is used (see Fig. 8.7). Each emoji is assigned a matching score of how well it matches the query word(s), based on the results of the semantic search, considering the number and source of occurrences, and the *Unicode* codepoint length (e.g. `U+1f474` is more specific than `U+1f474 U+1f3fb`). A score is assigned to each of the criteria, as follows:

1. Name (NV):
$$NV = \frac{\#matching\_words}{\#words\_name}, \tag{8.1}$$
where $\#matching\_words$ denotes the number of words in the name that match the word(s) searched and $\#words\_name$ is the total number of words composing the name.

2. Definition (DV):
$$DV = \frac{\#matching\_words}{\#words\_definition}, \tag{8.2}$$
where $\#matching\_words$ represents the number of words in the definition that match the word(s) searched and $\#words\_definition$ denotes the total number of words in emoji definition.

3. Keywords (KV):
$$KV = \frac{\#matching\_keywords}{\#keywords}, \tag{8.3}$$
where $\#matching\_keywords$ represents the number of keywords that match the word(s) searched and $\#keywords$ is the total number of keywords of the emoji.

4. Sense (SV):

$$SV = \frac{\#matching\_senses}{\#senses}, \tag{8.4}$$

where $\#matching\_senses$ is the number of senses that match the word(s) searched and $\#senses$ denotes the total number of senses of the emoji.

5. Unicode Codepoint (UV):

$$UV = \frac{1}{codepoint\_length}, \tag{8.5}$$

where $codepoint\_length$ represents the codepoint length.

The criteria are then weighted according to the following formula:

$$matching\_score = NV \times 0.3 + DV \times 0.15 + KV \times 0.3 + SV \times 0.2 + UV \times 0.05$$

The $matching\_score$ reflects how well the emoji matches the given word(s). The criteria have different weights due to the importance of each one (e.g. a word in the *name* is more important than in a *sense*). Moreover, *name*, *keywords* and *description* were initially gathered from the *Unicode Consortium*, whereas *senses* were based on user attribution and may be more ambiguous. After the search is concluded, a list of emoji sorted by $matching\_score$ is returned. This list can either be presented as existing emoji or used in the blending process as described in Section 8.1.2.2 (e.g. the *coffin* emoji for the word "go" in Fig. 8.5).

8.1.2.2 *Generation of visual representations* (T2)

In T2 the system behaves differently, depending on the number of introduced words. In the case of single-word concepts, the blending between emoji of the same word does not occur, e.g. two existing emoji for *car* (the red and orange in Fig. 8.5) are not blended together to represent the concept *car*. This would only happen if the concept introduced was "car car". Instead, the *Concept Extender* and the *Emoji Searcher* components are used to get emoji to blend (see Fig. 8.5).

The *Concept Extender* component is used to query *ConceptNet* for a given word, obtaining *related concepts*, sorted according to *ConceptNet* weight system. It is used in two situations: (i) when the user introduces single-word concepts to retrieve double-word concepts to use in the blending (e.g. for *car* the system obtains *go fast*, see Fig. 8.5); (ii) when the *Emoji Searcher* component does not find any emoji for a given word (e.g. *theory* does not have any matching emoji so the system uses the related concept *idea*, obtained with the *Concept Extender*, see Fig. 8.5).

In the case of introduced single-word concepts, we only consider double-word related concepts (e.g. *go fast* in Fig. 8.5) as initial experiments indicated that using emoji from two single-word related concepts would result in blends unrelated to the introduced concept. After obtaining the double-word related concepts, the *Emoji Searcher* component (already described for T1) searches for emoji for each word (e.g. in Fig. 8.5 the *coffin* emoji is obtained for *go*, and the *fast forward* for *fast*). These emoji are then used in the blending process.

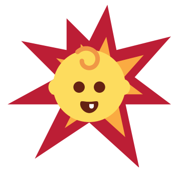


Figure 8.8: Blend produced for "generation".

On the other hand, when the user introduces a double-word concept, the system firstly searches for existing emoji for each word, using the *Emoji Searcher* component (already described). If emoji are found for both words (e.g. *wine glass* emoji for *wine* and *polo player* for *polo* in Fig. 8.5), a process of blend is conducted. If the system does not find existing emoji for both words, a search for related concepts is performed, using the *Concept Extender* component (already described). An example is shown in Fig. 8.5, in which no emoji is found for *theory*. The system uses the *Concept Extender* component to obtain related concepts (e.g. *idea*). After getting the related concepts, the system uses the *Emoji Searcher* to search for matching emoji (e.g. *light bulb*). If the search is successful, a blending process is conducted.

So far we have described how knowledge from the different resources is used to generate novel representations. One example is the blend for *generation* (Fig. 8.8). Firstly, the *Concept Extender* is used to retrieve the related concept *baby boom*. Then, semantic knowledge associated with emoji is used by the *Emoji Searcher* to obtain matching emoji: the *baby* is obtained through a match in the emoji's name; and the *collision* emoji is obtained through a match with the emoji's keyword "boom". The final step consists of a visual blending process, which makes use of *attribute-based* and *positioning* knowledge, retrieved from existing emoji (i.e. the *baby* emoji is placed according to the position of the *collision* emoji).

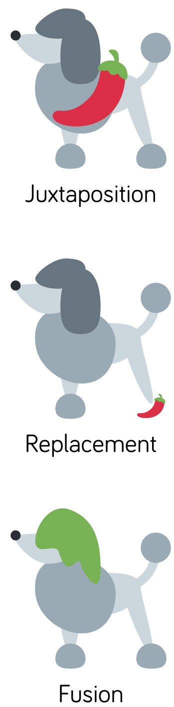


Figure 8.9: Types of blends

This process of visual blending occurs in the *Emoji Blender* component, which merges two input emoji (emoji A, the *base*, and emoji B, the *replacement/modifier*). The emoji to use are selected from the lists provided by the *Emoji Searcher*. In practice, the system uses two emoji, which are retrieved based on two words (either the *initial input* or a result of a *conceptual extension*, as explained before). The system uses the emoji of the second word as a *base* and the emoji of the first word as a *replacement* (see Fig. 8.5). In terms of blending, we consider three different methods (Fig. 8.9), based on the ones proposed by Phillips and McQuarrie (2004). The first method is *Juxtaposition* (JUX), in which the two emoji are put side by side or one over the other. Examples of JUX are the blends for *car* and *game theory* in Fig. 8.5. The second method is *Replacement* (REP), in which part of emoji A is replaced by emoji B. An example of REP is the blend for *wine polo* the water is replaced by wine (see Fig. 8.5). The third method is *Fusion*, in which the two emoji are merged together by exchanging parts. In this first iteration of the system, only JUX and REP were implemented.

8.1.3 *Presenting visual representations*

In the previous section, we presented the general architecture of *Emojinating*, describing its components and tasks. In this section, we explain how the results are shown to the user.

8.1.3.1 *Summary of the system's output*

In summary, the system takes a text *concept* (one or two words) and outputs a set of *visual presentations*, consisting of:

- existing emoji (E): emoji that have the exact text of the *query concept* in their semantic data. For example, if the user inputs "hot dog" the system uses the *Emoji Searcher* component to search for emoji that include exactly "hot dog" in their data, retrieving 🌭.
- related emoji (R): emoji that match a concept that is related to the queried one. The system uses the *Concept Extender* component to retrieve related concepts from *ConceptNet* and then uses the *Emoji Searcher* to search emoji that match the related concepts. The system is configured to search for both 1st degree (*queried concept* → *related concept* → *emoji*) and 2nd degree (*queried concept* → *related concept* → *related concept* → *emoji*) related concepts. For example, if the user inputs "freedom" the system will retrieve no existing emoji and the "statue of liberty" emoji 🗽 for 1st degree related concepts.
- emoji blends (B): blends that are produced by combining existing emoji that are related to the *queried concept*. For the producing blends, the system uses between two (if the user inputs a double-word concept that has existing emoji for each of its words) and three components (e.g. in the case of single-word concepts). In the case of single-word concepts, the blends are generated using double-word related concepts retrieved from *ConceptNet* using the *Concept Extender* component, as described in Section 8.1.2.2.

Figure 8.10: Visual blends for "rain man" using the same emoji. The top blend uses juxtaposition and the others use replacement.

The system's output is a variable number of visual representations for the introduced concept, composed of existing emoji (E), related emoji (R) and generated blends (B). The number of elements in each of these groups varies in quantity, depending on the data found.

The number of blends produced also depends on the nature of the query concept (single-word or double-word). The first iteration of the system has a deterministic nature, always showing the same results. For single-word concepts, it firstly extends the query concept to double-word related concepts and then, for each related concept, it produces blends. For the production of blends from double-word concepts, it either uses the initial double-word input concept or extends each word to related concepts, depending on the availability of emoji. In any

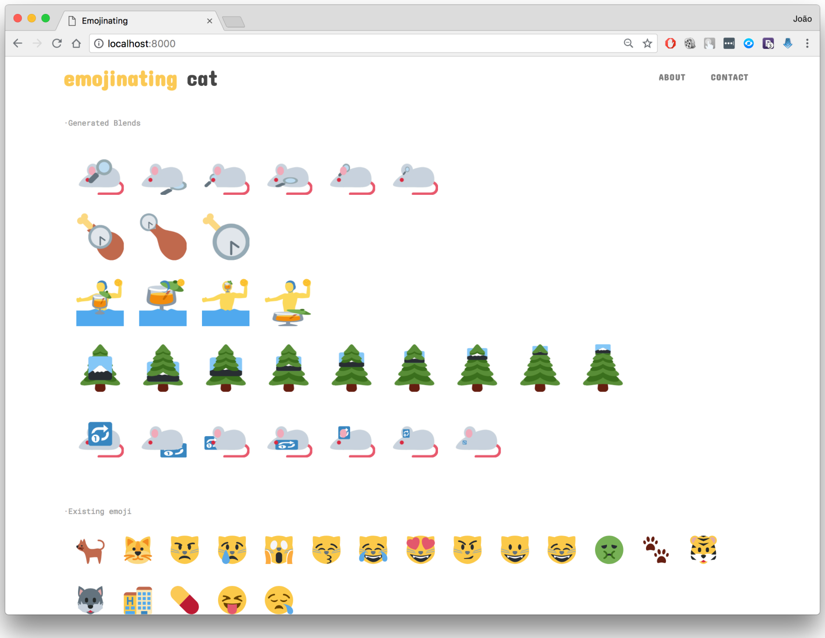


Figure 8.11: Interface of *Emojinating* version 1 showing blends and existing emoji for "cat".

case, for a given double-word concept (either input or related), the system only uses the emoji with the highest matching score for each word (two emoji in total).

For each emoji combination, the first iteration of the system produces the following blends:

- one JUX blend, in which the replacement emoji B is reduced in size and placed over the base emoji A (e.g. top blend in Fig. 8.10);
- REP blends for each possible layer replacement – in each blend, the replacement emoji replaces a different layer of the base emoji. In other words, a blend is produced for each part (layer) of the base emoji A: emoji B replaces the part using its position (e.g. in Fig. 8.10 the *rain cloud* emoji replaces the *moustache*, the *face shape*, the *hair* and the *nose*). All possibilities are generated.

8.1.3.2 *Interface*

The aim of the *Emojinating* is to allow the user to input a concept and receive visual representations of it. As such, a web-based interface was developed (see Fig. 8.11), being composed of two main areas: the *search area* and the *results area*. The *search area* contains a search field, in which the user writes words to search. After conducting the search and generation of emoji, the results are presented to the user in the *results area*. This area is divided into four sections: (i) the generated blends section, which shows the blended emoji; (ii) the existing emoji section, which shows emoji retrieved from the search for the introduced word(s); (iii) the related emoji (1st level) section, which shows emoji for directly related concepts to the one introduced; and (iv) the related emoji (2nd level) section, which shows emoji for indirectly related concepts (retrieved using related concepts).

8.2 EVALUATING PERFORMANCE WITH NGSL

In order to evaluate the system in terms of its capability to produce visual representations of concepts, we first focused on single-word concepts. Moreover, we were interested in assessing the importance of the several knowledge sources for emoji retrieval.

We begin by describing the setup of a test[2] for the assessment of the system's quality in terms of gathering existing emoji and generation of visual blends. Afterwards, we present and analyse the results. The test was conducted with the goal of studying the following aspects:

- the general capability of producing results;
- impact of the several sources of emoji semantic knowledge on the retrieval of emoji;
- quality of the results in terms of concept visual representation.

8.2.1 *Experiment Setup*

In order to evaluate the system, we decided to use it to produce visual representations for a list of concepts. We selected the New General Service List (NGSL) (Browne, 2014), as it consists of a core vocabulary of 2801 words for second language learners. We believe that using a core concept list is a good way to assess the system's performance in producing visual representations for concepts. As most emoji represent nouns, we decided to apply this restriction to the list. Using *RiTa*[3] library (part-of-speech tagging function), the list was reduced to a total of 1509 nouns.

The system was used to produce visual representations (existing emoji + blends) for each concept of the list and was then evaluated in terms of solution production and quality of concept representation.

8.2.2 *Results*

As previously mentioned, this experience was conducted to study three different aspects: (i) capability of producing results, (ii) impact of different semantic knowledge sources and (iii) quality of the results. We

2 The test was presented by Cunha, Martins, and Machado (2018a).
3 `rednoise.org/rita/`

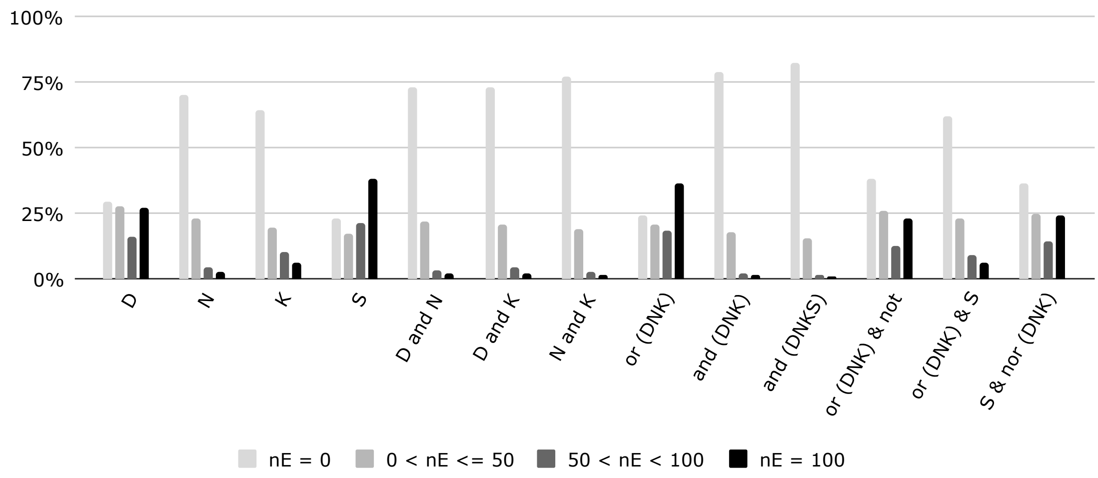


Figure 8.12: Sources of Semantic Information. Percentage of nouns in relation to the percentage of noun's existing emoji (*nE*) in terms of semantic information source, e.g. 29.45% of the nouns with existing emoji have none of their emoji ($nE = 0\%$) with *definition* as source of their semantic information (first bar on the left). Sources are: Definition, Name, Keywords and Senses.

divide the analysis of the results into two parts – the first focuses on points (i) and (ii), and the second on (iii).

8.2.2.1 *Production and Semantic Knowledge Sources*

In general, the system is able to retrieve visual representations that reflect the meaning of the noun, both related (e.g. "change") and blended (e.g. "roof") – see Fig. 8.13. From the 1509 input names, the system is only unable to produce results for four nouns ("protein", "incentive", "immigrant" and "refugee"), as observed in Table 8.1. It retrieves existing emoji for 927 nouns, 1st level related emoji for 1267 nouns, 2nd level related emoji for 1212 nouns and produces blends for 1043 nouns.

The most significant source of semantic information is *senses*, with 59.44% of nouns having the majority of their emoji related to *senses*, 38.3% have all the emoji (100%), and only 23.3% have none of the emoji (0%) related to *senses* (Fig. 8.12). It is also important to notice the value of *definition*, with 43.04% (26.86 + 16.18) of nouns with the majority of their emoji related to *definition*, 26.86% with all the emoji, and 29.45% with none of the emoji. These two sources highly contrast with the rest, as well as, with combinations among them.

8.2.2.2 *Quality of Visual Representations*

The system's ability to retrieve related emoji and produce blends does not mean that these correctly represent the concept. To assess how well the blends and related emoji represent each concept, we analysed them

Table 8.1: Results of producing representations for the NGSL dataset. Number of nouns with each type of emoji – related (R) 1st and 2nd level, blended (B), either R or B (R∪B), and neither R or B ($\overline{R}\cap\overline{B}$) – and the presence (E) or absence ($\overline{E}$) of existing emoji. The number of emoji considered in R does not include the ones that also exist in E.

		R 1ST	R 2ND	B	R∪B	$\overline{R}\cap\overline{B}$
E	927	853	683	707	921	6
$\overline{E}$	582	414	529	336	578	4
	1509	1267	1212	1043	1499	10

Table 8.2: Results in terms of quality of related emoji (R) and blends (B) – (1) none represents the noun, (2) bad, (3) neutral, (4) good and (5) obvious, expressed in number of nouns.

(a)

	1	2	3	4	5	
R	692	253	287	215	31	1478
B	668	187	108	74	6	1043

Table 8.3: Usage of generated visual representations (G), i.e. related and blends, and presence of existing emoji (E), expressed in number of nouns. Nouns with existing emoji were divided into: *good* (at least one existing emoji represents the noun) and *bad* (no existing emoji represents the noun). It shows the number of nouns in which one of the generated emoji was selected to represent the noun (S); and in which none was selected ($\overline{S}$).

(b)

		GS	G$\overline{S}$	$\overline{G}$	
E	*good*	112	675	4	791
	bad	65	69	2	136
$\overline{E}$		288	290	4	582
		465	1034	10	1509

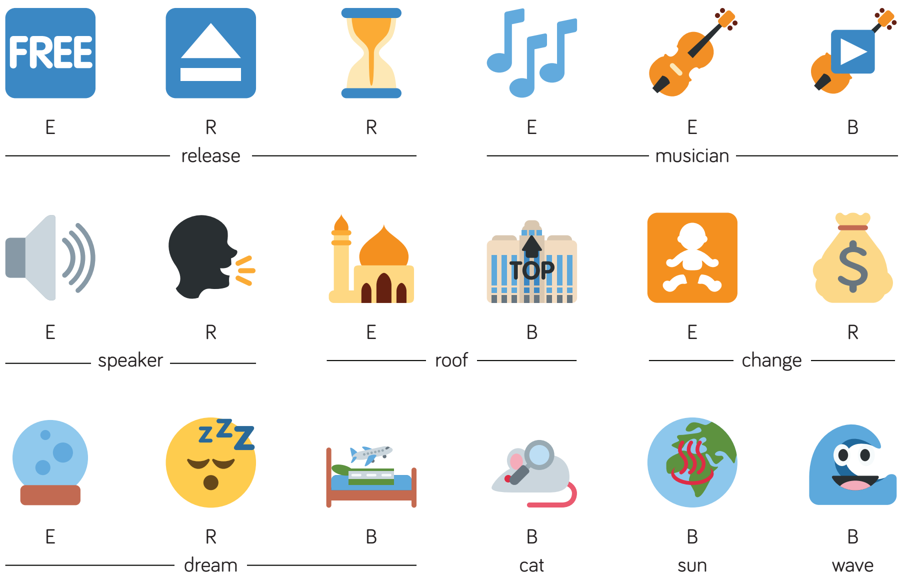


Figure 8.13: Examples of the system output for several concepts, showing existing emoji (E), related emoji (R) and produced blends (B).

and assigned a value from 1 (does not represent the noun) to 5 (represents in an obvious way) to each group (related emoji and blends). This assessment focuses on the best exemplar of each group (if one exists). The obtained results can be seen in Table 8.2.

On the other hand, some of the sources used to retrieve existing emoji are not official but result from user attribution (e.g. *senses*). For this reason, there is no guarantee that they represent the concept well. To evaluate the quality of the existing emoji we assigned a binary value corresponding to whether it represents the concept (*good*, e.g. 🔊 for *speaker* in Fig. 8.13) or not (*bad*, e.g. 👨‍🎤[4] for *actor* in Fig. 8.14). Afterwards, we identified if at least one of the generated emoji (related or blended)[5] can be selected to represent the noun (S) – i.e. it is as good or better than the existing emoji.

From this analysis it is possible to divide the nouns into different groups (see Table 8.3):

1. **GS∩E** – a generated emoji was selected to represent the noun (S) despite the presence of existing emoji (E). Three situations occur: **(a)** *Good* E but the generated ones are even better. This is the best case scenario and had an incidence of 112 out of 921 emoji with Existing and Generated emoji (12.16%), which we consider a good result – e.g. *musician*, *release* and *roof* in Fig. 8.13; **(b)** *Bad* E and the generated ones are better (e.g. Fig. 8.14, the existing emoji

4 Notice that 👨‍🎤 is the image of emoji *man singer* (`U+1F468 U+200D U+1F3A4`) in *Twemoji* 2.3, which was used in the experiment. Recent versions have a different image.

5 The term "generated emoji" is used for ease of writing to refer to related emoji and blends, even though related emoji are retrieved rather than "generated" and "blends" cannot be truly considered "emoji"

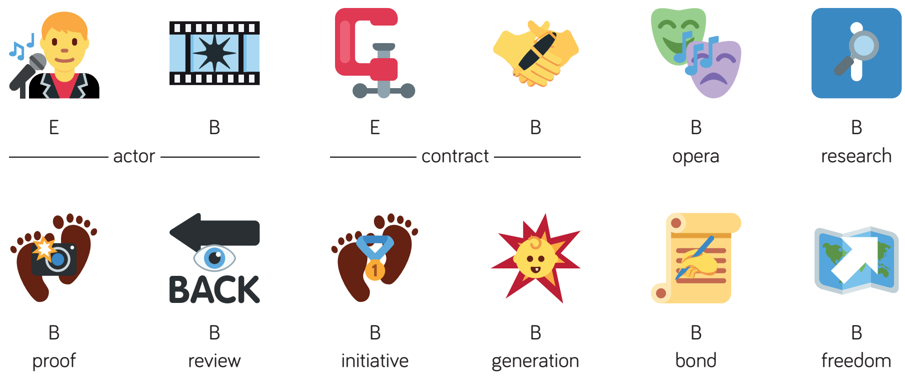


Figure 8.14: Examples of generated blends (B) for single-word concepts that either not have existing emoji, e.g. *freedom*, or have non-representative ones (E), e.g. *actor*.

for *actor* depicts a singer). We do not consider the results for this group very good as the generated were only selected in 65 from a total of 134 nouns with generated and bad existing emoji. One reason for this may be the abstract nature of nouns; **(c)** *Equally good*. The generated emoji are as good as the existing emoji. This is often related to different meanings for the same noun – e.g. *change* and *speaker* in Fig. 8.13;

2. **$GS \cap \bar{E}$:** There are no existing emoji and the system is able to generate emoji that represent the noun well – e.g. *initiative*, *proof*, *generation* and *review* in Fig. 8.14;

3. **$\overline{GS} \cap E$:** The system is not able to generate anything better than the existing emoji. Two situations occur: **(a)** *Good* E. This is the case with most incidence (675 nouns). This is easy to justify as some nouns have officially associated existing emoji – e.g. *cat* and *sun* in Fig. 8.13. The fact that the generated were not selected, does not mean their quality is not good – it was just not enough to surpass the existing emoji. This may be due to the non-literal nature of generated emoji; **(b)** *Bad* E. Despite the bad quality of existing emoji, the generated ones are not considered better. One reason for this may be the abstract nature of the nouns;

4. **$\overline{GS} \cap \bar{E}$**: the system does not produce anything good enough to represent the noun. This is the worst situation.

Despite stating that the number of nouns with existing emoji is 927 (Table 8.1), the number of nouns well-represented with existing emoji is only 791 (Table 8.3). The number of nouns for which the system is not able to produce an adequate emoji is 365 ((*bad* E $\cap$ $\overline{GS}$) + ($\bar{E} \cap \overline{GS}$) + (*bad* E $\cap$ $\bar{G}$) + ($\bar{E} \cap \bar{G}$)). This means that the system is able to present the user with concept-representative solutions for 1144 nouns out of 1509 (an increase of 44.63% when compared to the initially well-represented 791 nouns, i.e. achieving a higher conceptual coverage than the emoji set on its own) – see examples in Fig. 8.13. It is important to bear in mind that the initial number of well-represented nouns would be even lower if we did not consider the emoji retrieved using non-official semantic knowledge (from *EmojiNet*). An example is case of the noun *contract* (shown in Fig. 8.14) for which the system is only able to find an existing emoji, *clamp* 🗜 (`U+1F5DC`), which is retrieved using information from *senses*.

8.3 USER STUDY WITH DOUBLE-WORD CONCEPTS

After conducting an evaluation focused on single-word concepts (described in Section 8.2), we decided to shift our attention to double-word concepts. In order to assess the quality of the blend generation process, a user study[6] was conducted, which focused on three things: ability to represent concepts, quality of the blends and degree of surprise. In this evaluation, we were more interested in exploring the creative potential of the system than in obtaining results for concrete concepts. In this section, we present the results of the user study.

8.3.1 *Experiment Setup*

In order to assess the quality of the system in terms of visual blend production, we conducted a user study. Firstly, a list of ten concepts was produced. In this study, we focused on made-up concepts to explore the creative potential of the system.

The concepts to use in the study were randomly generated on the website *Title Generator*,[7] resulting in the following set: *Frozen Flower*, *Secrets in the Future*, *Serpent of the Year*, *Silent Snake*, *Storm of the Teacher*, *The Darkest Rose*, *The Flame of the Swords*, *The Laughing Blade*, *The Sexy Moon* and *The Sharp Silk*.

The study consisted in giving the concept list to participants, who would input the concepts into the system to produce visual blends and then answer a series of questions related to their experience with the system and quality of the output. As such, each participant received the concepts one by one, in a random order – this was done to minimise the biasing of the results. For each concept, the participants were asked to conduct the following tasks:

- T1 introduce the concept and generate the blends;
- T2 answer if there is a blend that represents the concept (yes or no);
- T3 evaluate quality of representation from 1 (very bad) to 5 (very good);

6 The study was presented by Cunha, Martins, and Machado (2018b).
7 `ruggenberg.nl/titels.html`

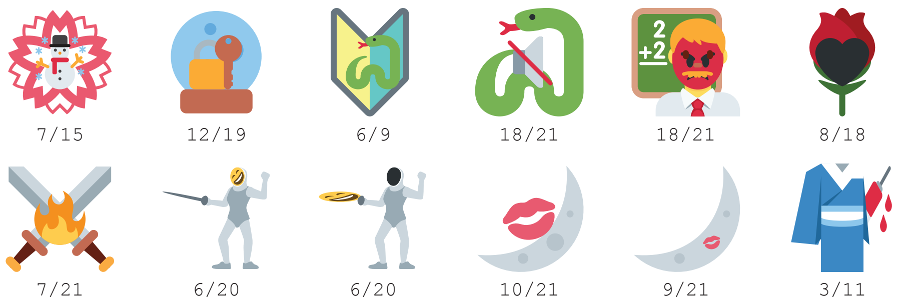


Figure 8.15: Blends selected as best representation for each concept (top 12). Below each blend is the number of participants who selected it and the total number of participants who selected a blend for that concept. The blends are ordered left-right, top-bottom, according to the order used in Table 8.4. Two blends are shown for *The Laughing Blade* and *The Sexy Moon*.

T4 identify degree of surprise from 1 (very low) to 5 (very high);

T5 select the best blend (only if a positive answer was given to T2).

A section for the participants to write optional comments was also included. Asking the user to select the best blend and then evaluate the system based on it may not be a common way of assessing a system's quality. However, in the case of our system, it serves the purpose as one of the goals of the system is to be used for ideation, in which having at least one good solution is enough.

8.3.2 *Results*

The study was conducted with 22 participants, who used the system to produce visual representations for all ten concepts. It is important to mention that the number of blends generated is variable and, consequently, the number of blends shown was not the same for every concept – e.g. for *Silent Snake* the system produces seven blends and for *Storm of the Teacher* 47.

The results obtained are shown in Tables 8.4 and 8.5. Overall, the system was able to generate blends that represented the concepts – 71.36% (157 out 220) of the answers to T2 were positive (Table 8.4) and the quality was above or equal to high (4) in 46.81% (103 out of 220) of the cases.

Moreover, the system is able to produce different blends that are considered interesting for the same concept. Figure 8.15 shows two blends for *The Laughing Blade* that were selected as the best by an equal number of participants. One reason for this may be the different interpretations for *The Laughing Blade*: the blade is literally laughing; or a non-literal interpretation related to a possible nickname of the swordsman, who laughs a lot.

Similarly, the best blend for *The Flame of the Swords* is literal and for *Storm of the Teacher* is metaphoric. The results concerning surprise seem

Table 8.4: Results of T2, T3 and T4, expressed in number of answers

	T2 (representation)		T3 (quality)		T4 (surprise)	
Concepts	Yes	No	<4	≥4	<4	≥4
Frozen Flower	13	9	6	7	7	6
Secrets in the Future	18	4	7	11	8	10
Serpent of the Year	5	17	1	4	1	4
Silent Snake	20	2	7	13	7	13
Storm of the Teacher	20	2	5	15	4	16
Darkest Rose	18	4	4	14	10	8
The Flame of the Swords	21	1	3	18	13	8
The Laughing Blade	16	6	12	4	7	9
The Sexy Moon	21	1	5	16	6	15
The Sharp Silk	5	17	4	1	2	3
Total	157	63	54	103	65	92
Mode			4		4	
Median			4		4	
Mean			3.73		3.59	

to reflect this difference: *The Flame of the Swords*, despite having a good quality score, was not considered surprising by the majority of the participants, whereas *Storm of the Teacher* was considered both surprising and of good quality.

The worst results were from *The Sharp Silk*, which was only considered concept-representative by 5 participants, from which only one assigned a quality score above or equal to high (4). Their opinion on the surprise criterion was also divided, resulting in two modes (2 and 5).

Most participants reported having difficulty in understanding some of the blends. Some did not recognise a shape (e.g. red shape of *Frozen Flower*), others had different interpretations (a planet instead of a crystal ball for *Secrets in the Future*) and others did not understand the reason behind a blend (e.g. *Serpent of the year*) – see Fig. 8.15. These were the main reasons for answering negatively to T2, and possibly for differences in the participants' opinion.

In the current implementation, only the emoji with the highest matching value is used. Changing this would increase the number of resulting visual blends and possibly lead to the generation of better ones, as the highest matching value does not necessarily result in the generation of the best blends.

8.4 SUMMARY

In this chapter, we presented and described *Emojinating* – a system that searches for *existing emoji* and automatically produces *emoji blends*,

Table 8.5: Mode, median and mean for T3 and T4 (only includes participants who answered positively to T2)

	T3 (quality)			**T4 (surprise)**		
Concepts	Mode	Median	Mean	Mode	Median	Mean
Frozen Flower	4	4	3.54	3	3	3.46
Secrets in the Future	4	4	3.67	4	4	3.50
Serpent of the Year	4	4	3.80	4	4	4.00
Silent Snake	4	4	3.80	4	4	3.80
Storm of the Teacher	4	4	3.95	4	4	3.85
The Darkest Rose	4	4	3.89	3	3	3.39
The Flame of the Swords	4	4	4.05	3	3	3.14
The Laughing Blade	3	3	3.19	3	4	3.63
The Sexy Moon	4	4	3.86	4	4	3.81
The Sharp Silk	3	3	2.60	2 and 5	4	3.60

based on a user-introduced word. To do so, it combines semantic network exploration with visual blending. In order to implement the system, we used the following resources: *Twitter*'s *Twemoji* dataset, *EmojiNet* dataset and *ConceptNet*.

In order to evaluate the system, we conducted two experiments. In the first experiment (Section 8.2), we assessed the system's quality in terms of output production, testing it with 1509 nouns from a list of core concepts. The system was able to produce results for the majority of the nouns, achieving good quality of representation and novelty.

In the second experiment (Section 8.3), we asked 22 participants to use the system to produce visual representations for ten concepts. The goal of the experiment was to assess three aspects: ability to represent concepts, quality of the blends and degree of surprise of the participants. The system was able to produce concept-representative visual blends and, for many cases, the participants stated that the results were different from what they were expecting.

As an ideation tool, this first version of the system has clear limitations. First, the system is *deterministic* and only shows results produced using the best emoji, not exploring the full search space. Second, the interface only allows the user to input the concept, which we consider a reduced interaction.

In a way, by running the core concept list on this deterministic version of the system we have established what we consider as the low limit in terms of solution quality, as the system does not explore the solution space to find better solutions nor is it prepared for automatic quality assessment – i.e. the focus is mainly on generating solutions, not on retrieving the best.

9 EMOJINATING: EVOLVING EMOJI BLENDS

In Chapter 8, we have presented the *Emojinating* system. It uses emoji as input visual representations for a process of visual blending in which they are combined with the goal of representing concepts given as input by a user. Despite being able to achieve a higher conceptual coverage than the one from emoji set (as described in Section 8.2.2.2), the implemented system does not employ an effective strategy for exploring the search space – it only considers the best emoji (semantic match) for blend generation. This approach ignores most of the search space and does not guarantee that the solutions are the most satisfactory for the user – one of the shortcomings identified.

In this chapter, we tackle the aforementioned issues by proposing an Interactive Evolutionary Computation (IEC) framework that combines a standard Evolutionary Algorithm (EA) with a method inspired by Estimation of Distribution Algorithms (EDAs) to evolve visual representations for concepts introduced by the user.

In order to assess the quality of the evolutionary approach and compare it with the deterministic version, we conducted two user studies. We present and discuss the experimental results obtained.

This chapter is based on the work presented in the paper by Cunha et al. (2019a).

9.1 CONTEXT AND RELATED WORK

In the domain of visual representations, computers have been made to draw on their own (e.g. McCorduck, 1991), used as *creativity support tools* for drawing (e.g. Lee, Zitnick, and Cohen, 2011) and even been given the role of *colleague in co-creative systems* (e.g. Davis et al., 2016). These examples, however, are more related to *Art* and shift away from the *Design* domain, in which a specific problem is addressed, e.g. how to design an icon to represent a given concept.

EAs are computational models inspired by the *Theory of Natural Selection* (Darwin, 1859) and *Mendelian genetics* (Mendel, 1865). They are normally used in problems in which it is possible to assess the quality of solutions based on a specific goal. However, the difficulty behind developing computational approaches to solve *Design* problems is that, in most cases, they greatly depend on human perception. For this reason, they can be seen as open-ended, as there is no optimal solution since they hinge on the user preferences. Thus, assessing quality is a complex problem on its own.

A possible way to tackle this problem is to develop systems that allow the user to choose which solutions are adequate. One of such approaches in the Evolutionary Computation (EC) domain is usually referred to as Interactive Evolutionary Computation (IEC). IEC is characterised by having a *user-centred* evaluation process, enabling human and machine to work together in the production of solutions. It has been considered suitable for open-ended design problems (Parmee, Abraham, and Machwe, 2008), since it is capable of accumulating user preferences and, at the same time, stimulating creativity.

IEC has been used in different domains such as *Fashion*, to produce shoe designs according to the preferences of the user (Lourenço et al., 2017); *Poster Design*, to evolve typographic posters (Rebelo and Fonseca, 2018); or even *Information Visualisation*, to explore the aesthetic domain (Maçãs, Lourenço, and Machado, 2018). In terms of symbol generation and visual representation of concepts, IEC has also been seen as a possible approach to solve problems.

Dorris et al. (2004) use IEC to evolve anthropomorphic symbols that represent different emotions (e.g. *anger*, *joy*, etc.). The genotype of each individual is a vector of nine real-valued numbers that correspond to the angles of the nine limbs (e.g. torso, left shoulder, right elbow, etc.).

Dozier et al. (2005) focus on emoticon design using an interactive distributed evolutionary algorithm – multiple processors working in parallel. It allows several participants to interact in simultaneously, evolving solutions that are the result of their judgements. The emoticons are represented as a vector of 11 integer variables, which corresponds to the *y-coordinates* of 11 points (e.g. the first three codified the left eyebrow).

Piper (2010) also uses a distributed approach, proposing an interactive genetic algorithm technique for designing safety warning symbols (e.g. *hot exhaust*). It uses previously drawn symbol components as input which were then combined to produce new symbols. According to Piper (2010), this distributed approach allows the replacement of the usual focus group in symbol design process with a group of participants interacting using computers in a network.

Hiroyasu et al. (2008) propose an interactive genetic algorithm that uses a crossover method based on probabilistic model-building for symbol evolution according to user preference. Each individual (symbol) is a combination of a colour (HSB model) and a shape (from a set of eight different shapes).

Estimation of Distribution Algorithms (EDAs) are based on the idea that statistical information about the search space can be extracted and used to modify the probability model, reducing the search space and leading faster to good solutions (Gong, Yan, and Zuo, 2010; Pelikan, Goldberg, and Lobo, 2002). Our approach takes inspiration from EDA methods to provide a way to quickly and efficiently search for solutions that match the user preferences.

9.2 IMPLEMENTATION

In Section 8.2.2.2, we assessed the system's performance using 1509 nouns from the New General Service List (NGSL) (Browne, 2014) and reported that the system was able to produce emoji for 75% of the list. Despite considering these results as good, the tested system only uses the best emoji (semantic match) for each concept, not exploring the full search space. With this in mind, we propose an evolutionary approach to explore the search space and find visual representations of concepts that match user preferences.

9.2.1 *Evolutionary Approach*

To implement this new version of *Emojinating*, we use the *Emoji Searcher* and the *Concept Extender* components described in Section 8.1.2, and we introduce a novel approach to explore the search space using IEC techniques. As such, an evolutionary system was implemented. The *Emoji Blender* component was modified in order to work together with an evolutionary engine, in which the generated blends are the phenotype of individuals.

The approach has a two-level evolution: on a *macro level*, it uses a method that takes inspiration from EDAs to direct the search to areas that match the user preference; on a *micro* and more specific level, it uses a standard EA to focus the evolution on certain individuals. The approach is schematically represented in Fig. 9.1.

9.2.1.1 *Concept Tree and General Evolution*

The *Emoji Searcher* and *Concept Extender* components are used together to produce a graph-like structured object from the user introduced concept (T1 in Fig. 9.1). This object – which we refer to as *Concept Tree* (CT) – stores the conceptual and emoji data produced from the analysis of the concept (Fig. 9.2). It has two different levels: *concept level*, in which related concepts are connected (e.g. the concept *god* is connected with the related concepts *judge people*, *judge men*, *justify hate* and *quiet storm*); and the *emoji level*, which stores the sets of emoji retrieved for each concept (e.g. *judge* has a set of emoji and *men* has another, Fig. 9.2).

The complexity of the *Concept Tree* object depends on the type of concept introduced by the user. If it is a single-word concept (e.g. *god*), related double-word concepts are searched and then emoji are retrieved for each of the words. If the user introduces a double-word concept, no related concept is required and the system directly retrieves emoji.

Taking inspiration from EDA methods, a *weight value* is assigned to every concept in the set of related concepts (in case they exist) and to each emoji. These weights are also stored in the *Concept Tree* object. When we generate new individuals (T2 and T8), the weights are used

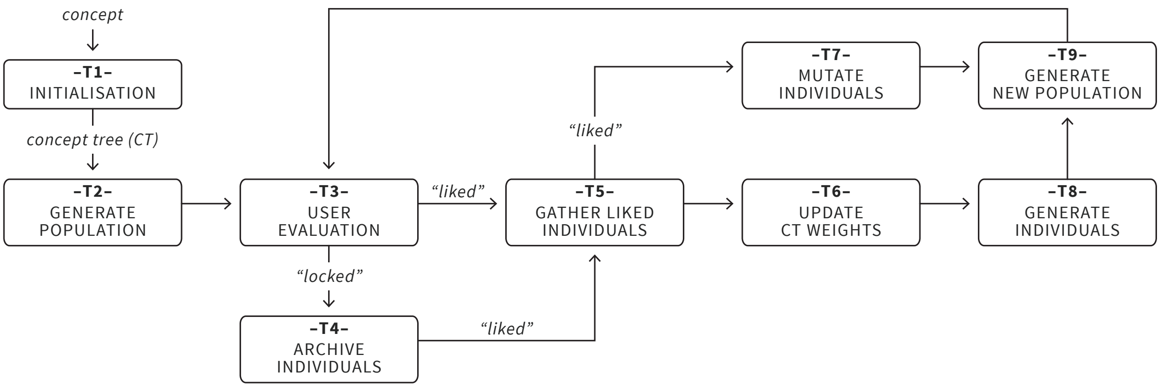


Figure 9.1: Evolutionary framework diagram, showing tasks (T1-9) and objects, e.g. *Concept Tree* (CT). The user starts by introducing a concept, which is used to generate a random set of solutions (T1 and T2). Then, the user evaluates the individuals by selecting the ones that fit their preferences (T3), referred to as "liked individuals". The user can also select the individuals to be stored in an archive (T4), referred to as "locked individuals". After the evaluation, the fittest individuals ("liked") are gathered from the population and from the archive (T5). The gathered individuals are then used to produce offspring through mutation (T7), and to update the weights of the *concept tree* (T6) – a graph-structured object with which new individuals are generated (T8). This process can be repeated indefinitely until the user is satisfied.

to select both the concept and the two emoji for the new individual – the higher the weight, the greater chances it has of being selected. Initially, the weights are all set to 1 and are updated in each generation according to user preferences (T6 in Fig. 9.1).

9.2.1.2 *Representation*

The emoji from *Twitter*'s *Twemoji* dataset are composed of layers – e.g. the *storm* emoji in Fig. 9.2 has two layers, a *cloud* and a *lightning*. As already mentioned, each visual blend is the phenotype of an individual. The individuals are encoded using a two chromosome genotype, which codifies the combination between two emoji. The first chromosome (*c1* in Fig 9.2) stores the two emoji (*e1* and *e2*) used in the blend. The second chromosome (*c2* in Fig 9.2) is responsible for defining the exchanges of parts that occur in the blend. This is codified by having each exchange stored as a gene (e.g. *g1*). Each gene corresponds to a set of two values: the first defines the part from *e1* that will be used as a *replacement* (-1 in Fig 9.2) and the second corresponds to the layer that will be *replaced* in *e2* (0.2 in Fig 9.2). As the number of layers is not the same among emoji, we use numbers in the [0,1] interval, which correspond to the position of the layer in the layer array. The value -1 can also be used, when the whole emoji is to be used instead of a layer (e.g. when a *juxtaposition* blend occurs). For example, for individual #2 in Fig 9.2 the whole *church* emoji is used as a *replacement* (encoded by the

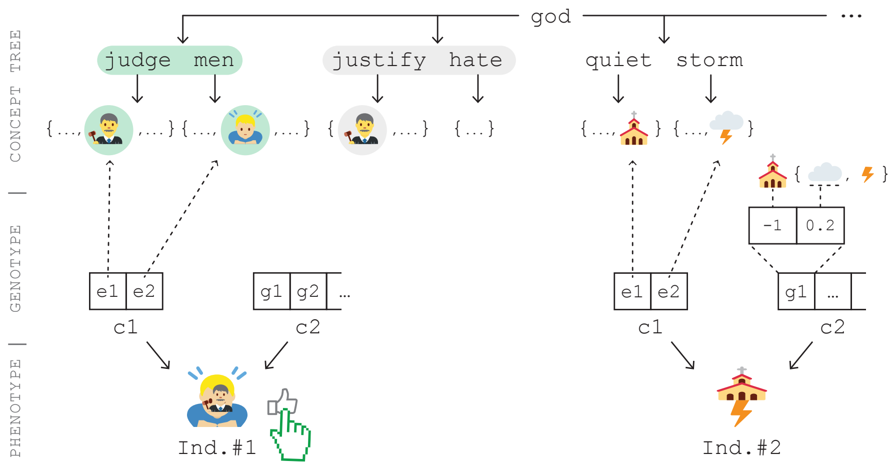


Figure 9.2: Individual representation and weight update system. The figure shows the two chromosomes (*c1* and *c2*) of two individuals' genotypes. It also shows gene 1 (*g1*) of *c2* from individual #2 in detail. Regarding the weight system, individual #1 is being "liked", which directly increases the weights of the concepts / emoji marked in green and indirectly increases the ones of concepts / emoji marked in grey.

-1 of *g1*) and the *cloud* layer is defined as a *replaceable part* (encoded by the 0.2 of *g1*).

In the implementation, two types of blends are used: *replacement* (a part of *e2* is replaced by *e1*) and *juxtaposition* (*e1* and *e2* are put side by side or one over the other). In both cases, only one exchange is encoded per individual. Although *fusion* blend type is not implemented in this version of the system, the chosen representation is already suitable for its integration.

9.2.1.3 *User evaluation*

In each generation, a population of 20 individuals is presented to the user, who is able to perform two different actions that affect the evolutionary process: mark individuals as "liked", which increases their fitness; and store them in the archive.

When an individual is "liked" (e.g. Ind. #1 in Fig. 9.2), the weights of the *Concept Tree* are updated (T6). It directly increases the weight of the related concept behind the individual and of the used emoji in the sets belonging to the concept (marked in green in Fig. 9.2). A process of indirect weight assignment is also used: the system searches for the emoji in other concept's sets and, if found, increases the weight of the emoji and corresponding concept (marked in grey in Fig. 9.2). This fosters related concepts that also use the same emoji and allows the system to find solutions that might also be of interest to the user. The weight increment is calculated based on the sum of weights of a set and it varies according to the type – a direct increment is 5% of the weight sum and an indirect is 2%. In order to make the evolutionary system work, the user does not need to classify every single candidate solution but only select the ones considered interesting.

9.2.1.4 *Weight equalisation*

A method of weight equalisation was implemented, which means that, as the evolutionary process progresses, there is a tendency towards an equal distribution of weights. The weights between concepts and between emoji inside sets will eventually converge to the same value, if not stimulated. This allows the system to achieve diversity even in late generations, avoiding unwanted convergence.

First, the average of weights inside a set is calculated. The weights that are above average are updated according to the following equation ($\text{EQ\_RATE} = \frac{1}{3}$):

$$new\_weight = (current\_weight - average\_weight) \times \text{EQ\_RATE}. \quad (9.1)$$

The method of weight equalisation is particularly useful when used together with an archive, which allows the user to increase population diversity and explore multiple areas of the search space, without losing individuals already considered good. Despite being different from classic EAs (in which convergence is a goal), this approach fits the problem, as the goal is to help the user find the highest number of interesting, yet considerably different, solutions.

9.2.1.5 *Archive*

An archive is often used to avoid the loss of population diversity by storing good individuals that can later be reintroduced in the population (e.g. Liapis, Yannakakis, and Togelius, 2013; Lourenço et al., 2017). Another possible use is to store the fittest individuals in order to use them to automatically guide the search towards unseen solutions (e.g. Vinhas et al., 2016).

In our case, diversification of the population is achieved with weight equalisation (as already described). Our archive works as a storage of individuals and has two main functionalities: (i) to save individuals and to avoid losing them in the evolutionary process; (ii) allowing the user to activate a permanent "liked" status that leads to a constant fostering of individuals. This option helps in situations in which the user has found an individual with an interesting trait (e.g. the use of a specific emoji) and wants to constantly stimulate it without having to manually do it on each generation. As explained before, selecting an individual as "liked" not only fosters its specific evolution but also has an effect on general evolution – changing *Concept Tree* weights and consequently affecting the generation of new individuals.

Moreover, storing the individual in the archive is the only way of guaranteeing that it is not lost when moving to the next generation. It allows the user to store solutions and focus on other possibilities while being able, at any time, to further evolve the stored individual, by activating the "liked" option. Combined with the weight equalisation, this makes it possible for the system to increase its diversity and, at the same time, avoid the loss of good individuals. In other approaches, such as (Lourenço et al., 2017), the archive is used to avoid the loss of population diversity by reintroducing stored individuals in the population. Differently, our strategy is to use the archive to allow the user to continuously change its exploration goal and try to find new promising areas in the search space.

9.2.1.6 *Mutation*

In addition to being used to update the *Concept Tree* weights (T6), user-evaluated individuals (the "liked" ones) are also employed in the production of offspring in each generation (T7). These are gathered from both the current population and the archive. From each "liked" individual, a set of four new individuals are produced (e.g. in Fig. 9.3, the four "bread-rhinos" in the population were generated by mutating the "liked" one in the archive). The parent individual goes through a mutation process, in which three types of mutation may occur: (i) *emoji mutation* (20% probability of occurring) – the emoji used as replacement is changed; (ii) *layer mutation* (80% probability of occurring per gene) – the replaced layer is changed (e.g. all "bread-rhinos" in the population except the first); and (iii) *blend type mutation* (5% probability) – this mutation changes the type of blend to *juxtaposition*, in which the emoji are used together and no replacement occurs (e.g. the first "bread-rhino" in the population). If a blend type mutation happens, no layer mutation occurs. The values presented were empirically obtained through experimentation and adjustments.

The use of the *layer* and *emoji mutation* types covers two situations: (i) adequate emoji are being used but the layer is not the correct; (ii) the exchange of layers is considered good but using different emoji may lead to a better solution.

9.2.1.7 *Offspring*

The offspring produced from parent individuals (T7) are added to a pool, from which they are afterwards randomly selected for the next generation. The number of individuals in the population is constant (20). As such, there is a maximum percentage of the new population (30%) that is used for individuals generated from parents through mutation. The remaining percentage corresponds to new individuals generated from the *Concept Tree* (T8). When generating individuals from scratch using *Concept Tree*, the probability of *juxtaposition* is set to 20% and of *replacement* to the remaining 80% – *replacement* can lead to many

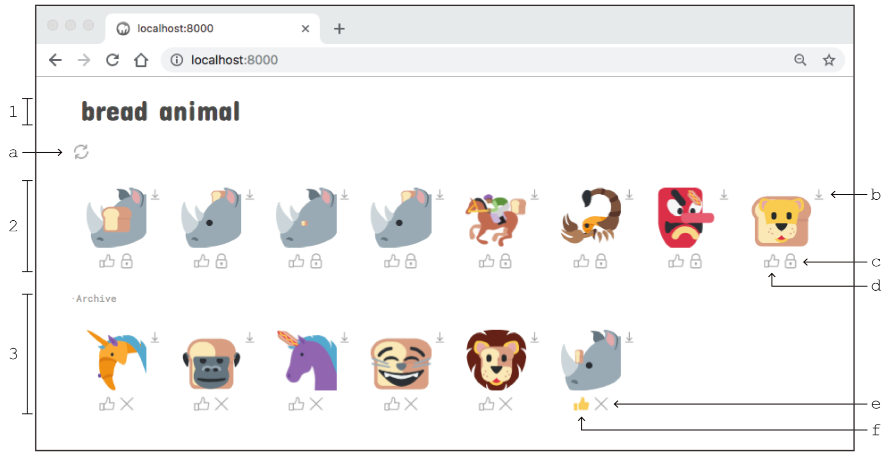


Figure 9.3: Interface of *Emojinating* version 2. The interface is divided into three areas: *search area* (1), *population area* (2) and *archive area* (3). There are five different button types that allow the user to interact with the system: *next generation* (a), *download* (b), *lock* (c), *like* (d) and *remove from archive* (e). A "liked" individual has an activated *like* button (f). The number of individuals in the population was intentionally reduced to increase legibility.

more different solutions than *juxtaposition* and, as such, it should occur more frequently.

9.2.2 *Interface*

The IEC system was implemented as a *web-based application*, which allows user interaction (see Fig. 9.3). The interface has three areas: the *search area*, the *population area* and the *archive area* (1–3 in Fig. 9.3). The *search area* is where the user introduces the concept (e.g. *bread animal* in Fig. 9.3). The *population area* presents the current population, showing the visual representation of the blends. Each individual has buttons: the *like*, which is used to evaluate the individual (*d* in Fig. 9.3); the *lock*, which stores the individual in the archive (*c* in Fig. 9.3); and one to download the visual representation of the individual (*b* in Fig. 9.3). Individuals in the *archive area* also have a *like* button, which is used to activate or deactivate the evaluation of the individual (the choice is maintained between generations), and a button to remove it from the archive (*e* in Fig. 9.3).

9.3 EVALUATION: DETERMINISTIC VS EVOLUTIONARY

To assess the quality of the evolutionary approach, we started by conducting a user study[1] to evaluate whether the evolutionary approach could lead to better solutions than the ones produced by the non-evolutionary deterministic version of the system (described in Chapter 8).

1 The study was presented by Cunha et al. (2019a).

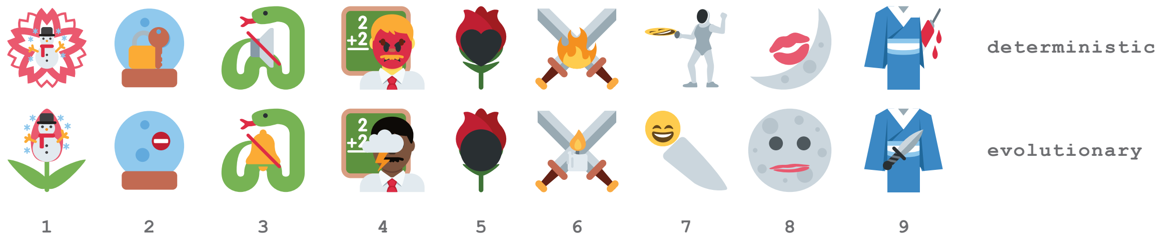


Figure 9.4: Blends used in user-survey for the concepts (1-9) *frozen flower*, *secrets in the future*, *silent snake*, *storm of the teacher*, *the darkest rose*, *the flame of swords*, *the laughing blade*, *the sexy moon* and *the sharp silk*. Blends in the top row were obtained with the deterministic version of the system and the ones in the bottom row with the evolutionary version.

9.3.1 *Experiment Setup*

In Section 8.3, we described a user study in which the system was used by 22 participants to generate visual representations for a set of ten concepts and the best solutions were collected. To compare the two approaches, we used our system to produce solutions for the same concepts (see Table 9.1) and conducted a survey with participants to assess if our system produced better solutions.

The survey was designed using the multiple-choice format, with one question for each concept, following the model:

Which of the following images represents better: [*insert concept*]*?*

Each question had four randomly ordered answer options: image produced with the evolutionary approach (referred to as "evolutionary image"), image obtained with the deterministic approach (referred to as "deterministic image"), "equally good" and "none".

In order to produce the survey, we used the system to select a good representation for each concept (see Fig. 9.4). Despite the risk of introducing bias towards our preferences, this was a necessary step to reduce the number of options presented to the user. One of the concepts (*serpent of the year*) was not used because we were not able to find any good solution different from the one presented in Section 8.3.

9.3.2 *Results*

The survey was conducted with 31 participants, with ages between 18-32. The results are shown in Table 9.1. We can see that for two of the concepts (*frozen flower* and *the laughing blade*) the evolutionary image was selected as better representing the concept by the majority of the participants; and for *the darkest rose*, 25.8% selected it as better or "equally good". Moreover, for *the sharp silk*, despite the majority of the participants selecting the option "none" (consistent with previous results,

Table 9.1: User study results expressed in percentage for each concept.

		answers (%)			
#	**concept**	evolutionary image	"equally good"	deterministic image	"none"
1	*frozen flower*	54.8	12.9	16.1	16.1
2	*secrets in the future*	9.7	0	58.1	32.3
3	*silent snake*	12.9	22.6	61.3	3.2
4	*storm of the teacher*	22.6	9.7	58.1	9.7
5	*the darkest rose*	9.7	16.1	16.1	58.1
6	*the flame of swords*	0	6.5	90.3	3.2
7	*the laughing blade*	45.2	12.9	16.1	25.8
8	*the sexy moon*	19.4	0	64.5	16.1
9	*the sharp silk*	32.3	3.2	3.2	61.3

see Section 8.3), the evolutionary image still had better results than the deterministic one, which was only selected by one participant. All in all, our approach was competitive in four out of the ten concepts.

9.4 TESTING WITH CONCEPTS FROM NGSL

In order to further evaluate the evolutionary approach, we conducted a user study[2] in which we asked participants to use the system to produce visual representations for concepts from the NGSL.

9.4.1 *Experiment Setup*

In Section 8.2, we used a set of 1509 nouns from the NGSL (Browne, 2014) to evaluate the performance of the deterministic approach. It was assessed if each noun was represented by its (i) *existing emoji* and by its (ii) *related emoji* or *blends*. Based on the results of the study (see Section 8.2), we divided the noun list into four groups:

- *group 0:* the system was not able to produce blends. This group was excluded as it could not be used due to the lack of blends;
- *group 1 (g1):* system produces blends but neither the related emoji/blends nor existing emoji represented the concept;
- *group 2 (g2):* system produces blends and only the related emoji/blends were reported to represent the concept;
- *group 3 (g3):* system produces blends and the existing emoji were reported to represent the concept (the related emoji/blends may also represent).

2 The study was presented by Cunha et al. (2019a).

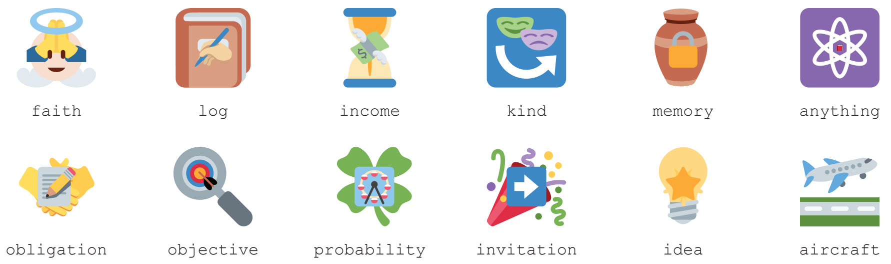


Figure 9.5: Examples of blends selected by the participants as good solutions.

Moreover, we crossed the list with a dataset of concreteness ratings (Brysbaert, Warriner, and Kuperman, 2014), obtaining a value of concreteness for each noun – from 1 (*abstract*, *language-based*) to 5 (*concrete*, *experience-based*). We divided each noun group in three subgroups to assess: (A) low concreteness, (B) medium concreteness and (C) high concreteness. These groups can be used to assess if there is any relation between *concreteness* and *representation easiness*.

We conducted a user survey in which each participant used the system to generate visual representations for nine randomly selected concepts (one from each subgroup). As the goal for this survey was to achieve maximum coverage of each subgroup, we decided to avoid noun repetition. Despite this, in low concreteness subgroups only few nouns existed – subgroup 1A had four nouns, 2A had three and 3A had five – which led to the repetition of nouns among participants for those subgroups. In total, 59 unique concepts were used. The participants used the system to evolve visual representations for the nouns, conducting only one run per noun and having a limit of 30 generations. They were asked to find individuals that represented the introduced noun and were allowed to stop the run before reaching 30 generations if they were already satisfied or if the system was not being able to further improve. For each noun, they were also requested to evaluate how well it was represented by the system, from 1 (very bad) to 5 (very good), and export the solutions that they considered the best among the ones that represented the noun (see Fig. 9.5).

9.4.2 *Results*

The survey was conducted with a total of eight participants. Table 9.2 presents the results obtained.

In terms of quality, the results show that the system is able to produce solutions with quality equal or above "good" for almost half of the concepts in groups 1 and 2 (10 out of 24), and for the majority of concepts in group 3 (15 out of 24). This is particularly important in group 1, for which the deterministic version was not able to find any satisfactory solution. Moreover, the participants were able to find more

Table 9.2: User study results for *quality*, *number of solutions*, *number of generations* and three combinations of quality (Q) / exported (E) /generations (G) that correspond to "early quit without results", "early quit with poor results" and "early satisfaction" (expressed in number of nouns and divided by noun group g1-g3).

	quality			# exported			# generations				Q≤3	Q≥4
	1	>1 ∧ ≤3	≥4	0	1	>1	<15	≥15 ∧ <30	30	E=0 ∧ G<20	∧ E>0 ∧ G<20	∧ E>0 ∧ G<20
g1	8	6	10	6	8	10	10	7	7	3	4	6
g2	9	5	10	6	11	7	8	10	6	4	2	7
g3	6	3	15	4	12	8	13	10	1	4	3	10

than one concept-representative solution in 34% of the runs (25 out of 72), e.g. *invitation* in Fig. 9.5.

We were able to compare the individuals selected as the best by each participant for each concept with the solutions obtained with the deterministic version of the system for the same concepts. In 38 out of 72 runs, the solution considered as the best was not produced by the deterministic approach. In addition, in 30 cases out of the 38 our solution was considered better than any of the solutions obtained with the deterministic version and in 5 was considered equally good. This shows that the evolutionary approach has clear advantages in comparison to the deterministic one.

Concerning the number of generations, in 80% of the runs (58 out of 72) the participants stopped before reaching the generation limit, which can be indicative of two things: the system could not produce blends that represented the concept or the user was already satisfied. To further analyse this matter, we used three combinations of *quality / exported / generations* that correspond to "early quit without results", "early quit with poor results" and "early satisfaction" (see Table 9.2). From the results, we can see that in 11 runs, the participant stopped without any exported solution before reaching 20 generations, which indicates that the system was not being successful. In addition, the column corresponding to "early quit with poor results" shows that in 9 runs the participant considered that the system would not get any better. On the other hand, in 30% of the runs (23 out of 72) the participant was satisfied before reaching the 20th generation, which means that the system was able to quickly evolve solutions that pleased the user.

One of the problems in IEC approaches is the weariness of the user (Lourenço et al., 2017). At the end of the survey, the participants evaluated the weariness degree of the task from 1 (very low) to 5 (very high) and 50% of participants rated it as very low in weariness and the other 50% as low. We also asked the participants to evaluate the surprise degree from 1 (very low) to 5 (very high) – 25% rated it as 3 and

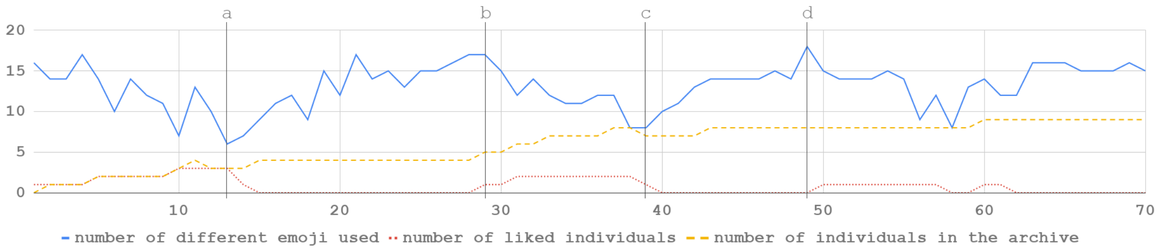


Figure 9.6: Metrics progression along the generations of a run for the concept *cell fish* (best viewed in colour). A video of the run can be seen at `rebrand.ly/evo19-thesis`.

75% as 4. This shows that the system is able to generate solutions that are unexpected.

When analysing the results, we could not observe any obvious relation between *concept concreteness* and *easiness of representation*. Our initial expectation was that concrete concepts would be easier to represent. The fact that we could not observe such correlation may indicate that using emoji blending to represent concrete concepts (e.g. *brain*) might not be the best approach. Moreover, some of the participants commented that they were trying to isolate an emoji, which is observed in some of the selected solutions – they tend to mostly show only one of the emoji (see blends for the concepts *anything* and *aircraft* in Fig. 9.5). However, further research is required on this subject as our remarks are only speculative and not statistically proven.

Another subject concerns the methods used in blend production. For single word concepts, the system gathers related double-word concepts to use in the blending process. The emoji belonging to each of the related concepts are not transferable to other concepts unless they are also in the emoji list of the concept, i.e. two individuals produced from different related concepts, one produced using *emoji* A and B and the other produced using *emoji* C and D, may never lead to the generation of an individual from *emoji* A and D. This is the reason behind some of the users commenting that they were not being successful in "combining" emoji from two individuals.

9.5 GENERAL ANALYSIS

The main goal behind a system for the visual representation of concepts is to be able to produce at least one good solution. Our evolutionary system allows the user to explore different and possibly unrelated areas of interest in the search space, often leading to several distinct solutions of good quality for the same concept.

To give an example of how the system reacts to user interaction, we show the progression of several metrics during one run (Fig. 9.6). It can be observed that the number of different emoji tends to decrease when solutions are marked as "liked", which shows that the population evolves towards similar solutions (e.g. from *b* to *c* in Fig. 9.6). The opposite is also verified: when no individual is "liked", the variation of the population tends to increase (e.g. from *a* to *b* and from *c* to *d* in Fig. 9.6). The increase in the number of individuals in the archive highlights its usefulness for search space exploration and reflects the capability of the system to evolve solutions that match user preferences.

In the cases in which the system was reported to not being able to generate anything that represented the concept, the reason was related to the gathering of semantic knowledge and not with the evolutionary engine. In general, the efficiency of the system is highly dependent on the existing semantic knowledge, emoji found and user perception.

Even though we consider that this version of the system has a good performance in terms of concept visual representation, there are certain aspects that should be addressed to improve the system.

Semantic weights The evolutionary engine does not take into account the semantic value assigned to related concepts and to emoji by the *Emoji Searcher* and *Concept Extender* components, which is used in the previous version to get emoji more related to the introduced concept. Considering these semantic values in the initial weight calculation may increase the fitness of the population in the first generation.

Blending and Mutation types The system currently uses two types of blending (*juxtaposition* and *replacement*) and three types of mutation (*emoji*, *layer* and *blend*). One way of improving the system could be to improve how the mutation operators work and to implement the *fusion* blend type.

Automatic fitness As already mentioned, one of the problems of IEC approaches is related to user weariness. Despite the participants reporting a low level of weariness, one of the possible directions for future work is implementing methods of automatic fitness attribution.

9.6 SUMMARY

In this chapter, we described the implementation of an *evolutionary* version of *Emojinating*, which combines a standard EA with a method inspired by EDAs. This approach allows the system to perform both a general evolution to direct the search to areas that match user preference and a focused evolution based on user-selected individuals. In order to do this, we used an archive to store individuals and selectively enable or disable their evolution.

We conducted two user studies to assess the quality of the evolutionary approach and compare it with the previous version of the system, which has a *deterministic* nature. The results show that the evolutionary approach allows the exploration of more of the search space and is able to present the user with better solutions. We also identified aspects that should be addressed to improve the system: (i) taking into account the semantic value attributed to related concepts and to emoji by the *Emoji Searcher* and *Concept Extender* components in the initialisation of weights, which may increase the fitness of the population in the first generation; (ii) considering *fusion* blend type; and (iii) improving *mutation operators*.

10

EMOJI VISUAL BLENDING FOR REPRESENTATION OF CONCEPTS

In the previous chapters, we have presented two versions of *Emojinating* – a *deterministic version* (Chapter 8) and an *evolutionary version* (Chapter 9). In addition to describing the implementation of the two approaches, we conducted a series of evaluation experiments. However, further experimentation is necessary to better study the suitability of emoji blending for concept representation purposes.

In this chapter, we start by describing the shortcomings of our previous studies. Then, we present the results of a preliminary test focused on production effectiveness (Section 10.3), in which the system was used to automatically generate visual representations for the concepts of two lists (*single-word* and *double-word*). Lastly, we describe a study that we conducted in order to compare the system performance in terms of *output quality* between single-word concepts and double-word ones (Section 10.4). We combine a more general analysis with one using statistical tools.

This chapter is based on the work presented in the paper by Cunha et al. (2020b).

10.1 CONTEXT

In the previous chapters, we presented the results of several studies, which were used to evaluate the system. Regardless of the obtained results, we consider that our analysis of emoji blending for concept representation is still narrow and has some shortcomings. We briefly analyse the limitations of our previous studies and identify the experimental questions of the study described in this chapter.

10.1.1 *Limitations of Previous Studies*

First, the conducted studies have limitations in several aspects. In the study described in Section 8.2, we used a list of *core single-word concepts* – the New General Service List (NGSL) (Browne, 2014) – and obtained promising results in terms of concept representation. Despite this, the analysis was conducted without resorting to external evaluators. Moreover, the study mostly focused on the capabilities of the system in terms of production effectiveness – i.e. whether the system is able to produce results – and on the origin of the semantic knowledge behind the emoji gathered by the *Emoji Searcher* component, giving little attention to the performance in terms of representation quality. To improve this, further testing with NGSL should be performed.

In the study described in Section 8.3, *invented double-word concepts* were used to test the system. Even though made up concepts are useful to assess quality for more creative purposes, the results obtained with them cannot be used to drawn conclusions for the representation of non-invented concepts. Then, in Section 9.3, we presented a study that used the same invented concepts but it was focused on comparing the *deterministic approach* with the *evolutionary one*.

In the study described in Section 9.4, the performance of the system with the NGSL was further tested by conducting a survey with eight participants, in which each generated blends for nine concepts. A total of 59 unique concepts were analysed, which we still consider a reduced number. Moreover, the analysis mostly focused on the performance of the evolutionary approach, overlooking questions regarding the blends (e.g. emoji used).

Second, one of the biggest shortcomings is the fact that, aside from the randomly generated concepts used in the studies described in Sections 8.3 and 9.3, only single-word concepts were tested. For this reason, one of our main goals is to test the system with a double-word concept list and compare the results with single-word ones.

Third, there are topics related to concept representation that would also be worth studying, one of which concerns *semantic concreteness* and its relation with the *easiness of visual representation*. So far, semantic concreteness was only used in the study described in Section 9.4 and mainly served to create groups of concepts for the survey (*low*, *medium* and *high* concreteness). Based on the results obtained, no obvious relation between concept concreteness and easiness of representation was observed. These results were not statistically proven and further research is required regarding semantic concreteness.

For the aforementioned reasons, in this chapter we conduct further user-testing with the NGSL core concepts, extend our analysis to a list of double-word concepts, and give more attention to the impact of semantic concreteness in visual representation of concepts.

10.1.2 *Experimental Questions*

One of our initial expectations was that concrete concepts should be easier to represent. Based on the fact that no correlation was observed in the study described in Section 9.4, we assume that using visual blending to represent concrete concepts may not be the best approach, which is in line with statements by some of the participants who admitted to purposely isolating one of the emoji used in the blend. This leads us to our first experimental question:

Q1: *Does the concreteness of concepts affect the performance of the system in terms of representation quality?*

Another question, which was already pointed out, is that little testing has been conducted with double-word concepts. As our visual blending approach uses double-word concepts as a starting point to gather emoji, we guess that the results for double-word concepts should be better in comparison to single-word ones. The reason behind our opinion is that for single-word concepts the system resorts to double-word related concepts (see Section 8.1.2.2) and, as such, the visual output should be more distant from the initial concept. Therefore, our second experimental question is:

Q2: *Is there a difference in the performance of the system between single-word and double-word concepts?*

On another topic, the emoji system is composed of visual representations which vary in terms of degree of abstraction – ranging from very pictorial emoji (e.g. *dolphin* emoji) to more abstract ones (e.g. a *triangle*). As the system purpose is to visually represent concepts using emoji, there is a question that should be addressed:

Q3: *Is there any connection between the concreteness value of the concepts and the visual concreteness of the emoji gathered by the system to represent it?*

These experimental questions are aligned with research question A of this thesis:

> *Can Computational Approaches, in particular those based on Visual Blending, be used for the visual representation of concepts?* (research question A described on p. 3)

10.2 EXPERIMENTATION RESOURCES

In order to address the experimental questions presented in Section 10.1.2, we first focused on producing resources that were necessary for the experimentation.

10.2.1 *Improved version of the system*

As mentioned in Section 9.5, one of the limitations of the interactive evolutionary approach presented was that the weight values assigned by the *Emoji Searcher* component were not being considered in the first generation of individuals, which meant that the best emoji were not being used. We produced an improved version of the evolutionary system, in which this issue was corrected.

10.2.2 *Concept lists*

Two concept lists were required: *single-word* and *double-word* (an overview of the lists can be observed in Table 10.1). In order to be in accordance with the work previously developed (Cunha et al., 2019a; Cunha, Martins, and Machado, 2018a), the single-word list to be used was the NGSL (Browne, 2014). The list had already been crossed with a concreteness ratings dataset (Brysbaert, Warriner, and Kuperman, 2014), with which we assigned a concreteness value to each noun – between 1 (abstract, language-based) and 5 (concrete, experience-based). Only two nouns could not be assigned a concreteness value – "criterion" and "dialog".

As such, only a double-word list was needed. To produce one, we extracted the double-word compounds from a noun-noun compound dataset (Fares, 2016), resulting in a list of 9612 compounds. Similarly to what was done with the NGSL dataset, we crossed the list with the dataset of concreteness values to obtain a value of concreteness for each word of the compounds, resulting in 4827 compounds with concreteness value for both words. As we noticed that some compounds used plural words, we converted them into singular words using the noms dataset from the Freeling multilingual language processing library (Padró and Stanilovsky, 2012). This increased the number of compounds with concreteness value for both words to 8065, which was considered as the final compound list. Out of the initial 9612 double-word compounds, only 1547 do not have concreteness for both words. Upon analysing them, we discovered that the lack of concreteness value was due to: (i) some of the words being compounds themselves (e.g. "new-product development", "fine-arts appraiser"); (ii) the use of the -ing form of a verb instead of a noun (e.g. "negotiating table", "organizing genius"); (iii) the use of non-words (e.g. "%"); and (iv) absence of the word in the concreteness dataset (e.g. "minimill", "ethylene" or "debenture"). We also experimented with word lemmatisation but it led to the change of meaning in some of the compounds, e.g. "controlling interest" was changed to "control interest" and "buying opportunity" to "buy opportunity". For the sake of maintaining the meaning, we discarded the lemmatisation approach.

The values of concreteness assigned to the individual nouns were afterwards used to calculate the average concreteness of the compound. However, it is important to bear in mind one aspect regarding this assignment of concreteness values: we do not consider the average concreteness of the nouns that compose the compound as the concreteness of the compound itself. The average concreteness is used to divide the concept list into groups – serving as a metric to guide our study – but should not be interpreted as the concreteness of the compound. Nevertheless, as the system produces blends for compounds using emoji retrieved for each word, we expect that the concreteness of the compound words may also play a role in how well the system performs. In an ideal situation, we would have the concreteness of the words and the concreteness of the compound but no dataset of compound concreteness values was available.

Table 10.1: Overview of the single-word (SW) and double-word (DW) datasets in terms of number of concepts and concreteness values. T is the total of concepts and "with C" is the number of concepts with concreteness score assigned. We divided the concepts with concreteness into groups, ranging from very abstract ($1 \le c < 2$) to very concrete($4 \le c \le 5$). The table also shows the median ($\tilde{x}$), mean ($\bar{x}$) and standard deviation (σ) of the concreteness scores.

	concepts						concreteness (c)		
	T	with C	$c \ge 1$, $c < 2$	$c \ge 2$, $c < 3$	$c \ge 3$, $c < 4$	$c \ge 4$, $c \le 5$	$\tilde{x}$	$\bar{x}$	σ
SW	1509	1507	100 6.64%	392 26.01%	447 29.66%	568 37.69%	3.57	3.531	0.994
DW	9612	8065	37 0.46%	1920 23.81%	4174 51.75%	1934 23.98%	3.485	3.493	0.647

An interesting result is the distribution of concepts in terms of concreteness (see Table 10.1): in single-word concepts, the majority of concepts is very concrete (37.69%) and in double-word, the absolute majority corresponds to concrete values (51.75%).

10.2.3 *Visual Concreteness*

In order to address Q3 (*concreteness of concepts* vs *visual concreteness of emoji*), we required a metrics of visual concreteness. Our approach was to use a categorisation of emoji in which the categories aligned with concreteness degree (e.g. *Animals* are more concrete than *Symbols*). Another goal was to be able to exploit the categorisation for visual blending purposes. However, most of the existing categorisations focus on thematic similarity (e.g. Donato and Paggio, 2017) and are not aligned with our needs. In this section, we briefly describe the development of an emoji categorisation with metrics of visual concreteness, which is presented in detail in Chapter 17.

Having the development of a blending-oriented categorisation as a target, we based our division on suitability to different blending techniques and, as such, the categories reflect similarities in visual representation (e.g. *perspective*). For example, animal faces are grouped separately from full bodies – a blending using fusion between two face emoji is much more coherent than between two objects with different configurations. With this in mind, we decided to use the following sequential criteria: (i) overall distinction between *entities*, *objects* and *places* – useful for visual blending, as the former two can be positioned in the latter (e.g. *dog* and *ball* positioned on a *beach*); (ii) grouping according to thematic – e.g. *Animals* (*entity*) or *Food* (*object*); (iii) grouping according to visual characteristics and similarity – e.g. *Faces-animal* are kept separate from *Animals*. The last criterion is mostly focused on the distinction between different blend types: *faces* are suitable for *fu-*

sion (exchange of parts), emoji from *Clothing-head* are suitable for *replacement* (replacing a *crown* with a *hat*) and from the more general *Clothing* (e.g. shoes) are more suitable for *juxtaposition*.

Focusing on visual concreteness, we use the term "visual concreteness" as comprising two dimensions (Prada et al., 2016): (i) *concreteness* – "stimuli that (...) refer to objects, materials or people should be considered concrete (...) otherwise, they should be considered as more abstract" – and (ii) *Meaningfulness* – "to what extent the stimulus conveys a meaning". Five independent evaluators assigned a visual concreteness value to each category, based on a general analysis of their emoji. The visual concreteness value ranges from 1 (very abstract, symbolic or ambiguous in meaning) to 5 (very concrete and with obvious meaning), e.g. *Drinks* has a visual concreteness value of 5. The categorisation was iteratively improved until there was full agreement among evaluators.

For more details on the emoji categorisation, we refer the reader to Chapter 17, where we focus on the development of the categorisation.

10.3 PRELIMINARY TEST: PRODUCTION EFFECTIVENESS

The first step was to assess how the system performs in terms of production effectiveness with each of the two different concept lists – *single-word* (NGSL dataset) and *double-word* (based on a noun-noun compound dataset). To do this, we followed the same approach used in the study described in Section 8.2: we used the system to automatically generate results for all the concepts in two lists.[1] The results are presented in Table 10.2.

Regarding the single-word concept list, the system was able to gather existing emoji for 927 concepts and generate blends for 1030. Concerning the double-word concept list, we initially used the original words, resulting in 5618 compounds with blends and 120 with existing emoji. Afterwards, we decided to use the list resultant from the process of changing plural words into singular ones (described in Section 10.2.2), which led to 6283 compounds with blends and 170 with existing emoji. The difference in blend number results from the loss of blends in 145 of the compounds, and blend gain in 810 – i.e. some of the compounds had blends when they had plural words but not when they were converted to singular, and vice versa.

When comparing the results from both lists, the percentage of concepts with blends is slightly higher in the double-word concepts – 68% of the single-word concepts and 77.9% of the double-word have blends.

1 This experiment was presented by Cunha et al. (2020b) (section Preliminary Test).

Table 10.2: Quantity of concepts (with blends and total) in each list, divided by concreteness value (c) from very abstract ($1 \leq c < 2$) to very concrete ($4 \leq c \leq 5$).

	single-word		**double-word**	
	w/ blend	*total*	*w/ blend*	*total*
$1 \leq c < 2$	45	100	23	37
$2 \leq c < 3$	241	392	1353	1920
$3 \leq c < 4$	292	447	3316	4174
$4 \leq c \leq 5$	452	568	1591	1934
total	1030	1507	6283	8065

10.4 EVALUATING OUTPUT QUALITY

The previous section focused on the performance of the system in solution production. However, the capability to retrieve emoji and produce blends does not necessarily mean that the system is able to correctly represent the concept. Therefore, an analysis in terms of output quality is necessary. We present an experiment,[2] in which we compare the performance of the system in terms of output quality, with single-word concepts from the NGSL (Browne, 2014) and double-word ones from the noun-noun compound dataset (Fares, 2016).

10.4.1 *Experiment Setup*

Our objective was to achieve maximum coverage of each of the two lists. For this reason, we decided to avoid concept repetition and established a goal of 100 concepts for each list. We divided each list into four groups based on the semantic concreteness metric (c): *very low concreteness* ($1 \leq c < 2$), *low concreteness* ($2 \leq c < 3$), *high concreteness* ($3 \leq c < 4$) and *very high concreteness* ($4 \leq c \leq 5$). Each participant would evaluate at least one concept from each group for both lists.

For each participant, the concepts were randomly selected assigning at least one concept from each group while still guaranteeing that the system would be able to produce blends. The participants used the system to generate results for each concept, conducting only one run per concept and having a limit of 30 generations. In each generation, a population of 20 individuals is shown to the user, who is able to interact with the system and select the ones that are best according to their preference. The participants were asked to evolve blends towards solutions that, in their opinion, represented the concept. They were allowed to stop the run before the 30th generation in two situations: (i) if they

2 This experiment was presented by Cunha et al. (2020b) (section Experiment #1).

were already satisfied or (ii) if they observed that the system was not improving anymore.

10.4.1.1 *Qualitative Data Collection*

For each concept the users were requested to conduct three tasks:

1. Classify how well the system represented the concept, using a scale from 1 (very bad) to 5 (very good);
2. Classify the surprise degree of the results, using a scale from 1 (very low) to 5 (very high);
3. Export the solutions that they considered the best, from the ones which they considered as concept-representative. In case no solution represents the concept, none should be exported.

10.4.1.2 *Quantitative Data Collection*

For each run, the following variables were automatically collected: the number of generations, the number of evaluated blends per generation, the number of blends in the archive per generation and the number of different emoji per generation.

10.4.2 *Results*

The experiment was conducted with a total of 22 participants, the majority of which tested a set of ten concepts; three could not reach this value due to time constraints. As such, each group was tested 27 times. From the single-word concept list, 108 different concepts were tested and no repetition occurred. In the case of the double-word concept list, due to a lack of concepts in the less concrete group, five concepts were tested twice, resulting in a total of 103 different concepts. These numbers surpass the initially established goal of 100 concepts for each list.

We analysed the data collected in the user studies using statistical tools. The data was inserted into the SPSS software (version 24). Initially, we performed an exploratory data analysis to understand the distribution of the data. We relied on the *Kolmogorov-Smirnov* test with a significance level of 0.05 to check if the data was normal. The test revealed that the data was not normal. To assess the role of *hiding* (i.e. only one of the emoji is identified in the blend) and *concreteness* on the *quality of representation*, the non-parametric multivariate ANOVA tests were performed. For the pairwise comparisons, we relied on the *Mann-Whitney non-parametric test*. All the tests used a significance level $\alpha = 0.05$.

For our analysis, we refer to the groups as s1-s4 in the case of single-word concepts (very low concreteness to very high concreteness, respectively) and d1-d4 for double-word concepts (very low concrete-

Table 10.3: User study results expressed in number of runs for *quality* (Q), *surprise* (S), two combinations of Q/S and *number of generations* (G). The results are divided into concreteness groups, ranging from 1 (very abstract) to 4 (very concrete), for single-word (S1-S4) and double-word (D1-D4). The table also shows the mode (*mo*), median ($\tilde{x}$) and standard deviation (σ) of quality and surprise.

	quality (Q)							**surprise** (S)								**generations** (G)		
	1	>1 ≤3	≥4	5	*mo*	$\tilde{x}$	σ	1	>1 ≤3	≥4	*mo*	$\tilde{x}$	σ	S≥4 Q≤2	S≥4 Q≥4	<15	≥15 <30	30
S1	7	11	9	1	4	3	1.27	1	6	20	4/5	4	0.98	9	5	14	8	5
S2	7	3	17	9	5	4	1.60	3	11	13	4	3	1.29	4	8	14	9	4
S3	4	11	12	4	3/4	3	1.27	3	12	12	4	3	1.20	3	5	11	14	2
S4	3	9	15	5	4	4	1.25	3	12	12	3	3	1.32	4	6	13	8	6
S	21	34	53	19	4	3	1.37	10	41	57	4	4	1.24	20	24	52	39	17
D1	6	6	15	9	5	4	1.61	1	7	19	4	4	0.93	6	12	17	10	0
D2	5	6	16	8	4/5	4	1.50	0	8	19	4	4	0.97	7	12	18	8	1
D3	4	11	12	5	4	3	1.35	4	14	9	3	3	1.16	2	5	19	8	0
D4	2	12	13	6	3	3	1.22	2	14	11	2/3	3	1.24	1	6	19	6	2
D	17	35	56	28	4/5	4	1.41	7	43	58	4	4	1.14	16	35	73	32	3

ness to very high concreteness, respectively). Based on the data presented in Tables 10.3 and 10.4, several observations are possible.

In terms of quality, no apparent differences between single and double-word concepts are observed. To further analyse this topic, we divided the study into two parts: a) analysis of single-word concepts, b) analysis of double-word concepts. Concerning the relationship between concreteness and the quality of the representation (Q1), we performed a pairwise analysis using the *Mann-Whitney* test and found out that there are no statistically meaningful differences for the two scenarios considered. Then we proceeded to analyse if the quality of the representation was influenced by the number of words that composed the concept. The results showed that there are no statistically significant differences between the representation quality for single-word or double-word concepts (Q2).

When using double-word concepts (noun-noun compounds) one aspect to take into consideration is the degree of *meaning compositionality* (as previously mentioned in Section 4.1.1) – the higher this value is, the greater the chances are that the meaning of the compound can be predicted from its parts – e.g. "shelf life" has a score of 0.196 and "growth rate" of 0.63 (Roberts and Egg, 2018). Considering that we are blending emoji retrieved for each of the words of the compound to visually represent it, it would be expected that the quality would depend on the concept's degree of compositionality. To assess this hypothesis, we used the dataset described by Roberts and Egg (2018). We first changed the plural words to singular and then crossed the dataset with our noun-noun compound list. From our list of 8065 double-word

Table 10.4: User study results expressed in number of runs for *number of solutions exported* (E), three combinations of *quality*(Q)/*exported*(E)/*generations*(G) that correspond to "early quit without results", "early quit with poor results" and "early satisfaction", and *number of different emoji* used in blends. The results are divided into concreteness groups, ranging from 1 (very abstract) to 4 (very concrete), for single-word (s1-s4) and double-word (d1-d4)

	exported (E)				Q≤3	Q≥4		**dif. emoji**	
				E=0	E>0	E>0		≥10	
	0	1	>1	G<20	G<20	G<20	<10	≤20	>20
s1	11	8	8	9	1	8	4	2	21
s2	6	11	10	4	2	12	6	3	18
s3	5	11	11	1	5	9	2	4	21
s4	5	9	13	1	3	10	2	1	24
s	27	39	42	15	11	39	14	10	84
d1	7	8	12	4	4	14	11	10	6
d2	5	17	5	2	4	16	13	8	6
d3	7	15	5	5	7	11	10	7	10
d4	3	14	10	2	8	12	10	9	8
d	22	54	32	13	23	53	44	34	30

concepts, only 4020 had a compositionality score. As this task was conducted after the tests had been done, we were only able to obtain a compositionality score for 29 out of the 103 different noun-noun compounds that had been tested – we had compositionality data for 32 runs as 3 concepts were tested twice. We calculated the correlation coefficient of the compositionality and the blend quality, obtaining a value of 0.265, which indicates that no correlation exists. Due to the reduced number of tested concepts with compositionality score, further studies are required on this subject.

Regarding *surprise*, based on the participants' comments there are two main situations for a high surprise: positive and negative. This is reflected in the results (Table 10.3), as for s1 almost half of the concepts with surprise higher than 4 was a negative surprise (quality equal or below to 2), which means that they were not expecting what they saw and they did not consider it as *concept-representative*. On the other hand, in s2 the majority of surprising concepts had also good quality. Interestingly, for abstract (low concreteness) double-word concepts (d1 and d2), the majority of the concepts with surprise above or equal to 4 had also quality equal or above 4.

The *exported* values in Table 10.4 reflect the number of blends that were exported by the participant, being considered concept-representative. The concepts are divided into three categories "none" (0), "one" (1) and "more than one" (>1). Focusing on s1 we observe that the majority of concepts in this group does not have any exported. Moreover, the quantity of concepts with exported >1 in s2 and s3 is twice as much as the ones of d2 and d3, which may indicate that for single-word con-

Table 10.5: Analysis of exported blends, organised in three levels: *concept*, *run* and *blend*. T indicates the total of concepts, runs and exported blends. For concepts, *no export* is the number of concepts without exported blends, *all hiding* refers to concepts in which all exported blends have emoji being hidden ("hiding") and *good blend* refers to concepts that have at least one blend without hiding. For runs, *with exported* refers to runs in which at least one blend was exported and *with hiding* refers to runs in which at least one blend had hiding. For exported blends, it is shown the number of blends with hidden emoji (*hiding*) and the number occurrences for each blend type: *juxtaposition* (JUX), *replacement* (REP) and *unidentifiable type* (?). The results are divided into concreteness groups, ranging from 1 (very abstract) to 4 (very concrete), for single-word (s1-s4) and double-word (d1-d4)

	concepts				**runs**			**exported blends**				
	T	*no export*	*all hiding*	*good blend*	*T*	*with exported*	*with hiding*	*T*	*hiding*	JUX	REP	*?*
s1	27	11	4	12	27	16	7	30	13	7	23	0
s2	27	6	6	15	27	21	10	41	16	6	33	2
s3	27	5	9	13	27	22	13	46	21	8	36	2
s4	27	5	10	12	27	22	16	45	25	11	34	0
s	108	27	29	52	108	81	46	162	75	32	126	4
d1	22	4	3	15	27	20	6	41	7	13	28	0
d2	27	5	5	17	27	22	5	30	6	19	11	0
d3	27	7	4	16	27	20	4	26	4	10	16	0
d4	27	3	2	22	27	24	2	37	2	19	18	0
d	103	19	14	70	108	86	17	134	19	61	73	0

cepts the user is more likely to find more than one satisfying solution (occurring in 38% of single-word concepts) than in double-word ones (29.6%). This may be justifiable if we observe the results for the quantity of different emoji.

For the single-word list, the absolute majority of concepts has the number of different emoji above 20 (Table 10.4). The same does not happen with double-word concepts, which may explain why not so many solutions were exported. This difference in terms of distinct emoji is easily justified by considering that for representing single-word concepts the system gathers several double-word related concepts (e.g. in Fig. 8.5 "go fast" was one of the related concepts used by the system for "car"), each with its own set of emoji, whereas for double-word concepts the emoji are only gathered for the query concept (e.g. "wine polo" in Fig. 8.5). This may also justify why there are more cases of early satisfaction ($Q \geq 4 \wedge E > 0 \wedge G < 20$) for double-word concepts (49% of the used concepts) than for single-word (36%) and also why the number of concepts to reach the 30th generation is much lower in double-word concepts than in single-word ones – less emoji thus less variability of solutions and need for exploration.

Regarding Q3 (*concreteness of concepts* vs *visual concreteness of emoji*), we started by analysing the single-word concepts. In specific, we were

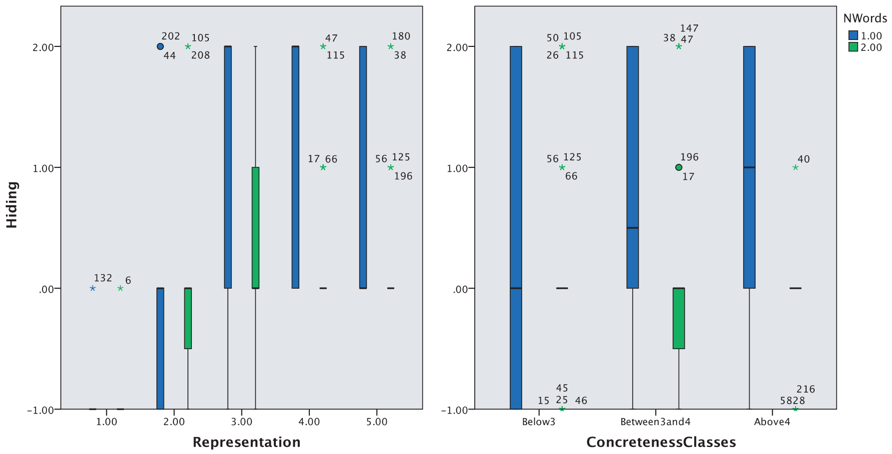


Figure 10.1: Box plots for single-word concepts (NWords=1) and double-word concepts (NWords=2) in terms of hiding – no exported blends (-1); exported and no hiding (0); exported and partial hiding (1); exported and obvious hiding (2) – versus representation quality (on the left), and versus concept concreteness (on the right).

interested in assessing if the concreteness of the words that composed the concept related to the concreteness of the emoji. The *Mann-Whitney* test revealed that there were no statistically significant differences between the concept's and emoji's visual concreteness in single-word concepts. After, we moved to the double-word concepts. In this study, our goal was to compare the concreteness of the individual two words with the concreteness of the emoji that were used in the blend, to see if there was any statistical difference between them. The results revealed that there are no differences between them. This means that *high concreteness* of the concept does not necessarily lead to a blend with emoji of *high visual concreteness*.

Then, we focused on the blends that were selected as most representative of the concepts (*exported ones*). We discovered that, in some cases, one of the emoji was partially or even totally hidden (see Table 10.5). Based on this finding, we concluded that this was most likely affecting our conclusions regarding *quality* (Q1 and Q2) because hiding one of the emoji is considered an exploitation of the system as it does not make use of visual blending. Moreover, we hypothesised that this was happening mostly in cases in which the concepts were composed of only one word. In order to further study this issue, we manually examined each exported blend and identified if there was an emoji being hidden and whether it was only partially or fully. To analyse this, we had to apply a multivariate analysis, since we had two grouping variables: *quality of the representation* and whether there was *hiding of emoji*. The results show that there are statistically significant differences in the results of the two lists in terms of quality when we consider the number of concepts with *hidden emoji* (p-value < 0.001887). These results can be visually confirmed by the box plots observed in Fig. 10.1. From the box plots, we can see that for double-word concepts there is barely any hiding. Also, the hiding mostly occurs when the concreteness is equal or above 3 (medium to very high) and when the representation quality is equal or above 3 (medium to very high). This means that a great number of concrete single-word concepts considered of good quality are actually exploiting the system and avoiding visual blending. In fact, this is particularly evident when we analyse the blends (see Table 10.5) – there were hidden emoji in 46 out of the 81 single-word runs that had exported blends, contrasting with the 17 cases in double-word ones. This is further highlighted when we consider that almost half of the exported blends for single-word concepts had hidden emoji (75 out of 162). Moreover, by observing the columns of the concept level we can see that for 29 out of the 108 single-word concepts all of the exported blends had hidden emoji. In contrast, for double-word concepts not only the number of concepts without exported blends was lower (19 vs 27) but also the number of concepts with at least one blend without hiding was higher (70 vs 52). These results indicate that our approach to visual blending is less useful in the representation of single-word concepts.

10.5 SUMMARY

In this section, we conducted a broad analysis of the performance of the system in concept representation through emoji blending by using the system with two lists of concepts: single-word and double-word. We presented and analyse the results obtained in (i) a preliminary test in which the system was used to generate results for all the concepts in two lists of concepts, and (ii) a user survey focused on the assessment of output quality.

We summarise our findings and answers to each of the experimental questions. Our findings are related to research question A of this thesis: *Can Computational Approaches, in particular those based on Visual Blending, be used for the visual representation of concepts?* (see p. 3).

Q1: *Does the concreteness of concepts affect the performance of the system in terms of representation quality?* No correlation between concreteness and the quality of the representation was initial found. However, we identified that hiding of emoji was occurring, which is an exploitation of the system as it does not make usage of visual blending. We concluded that hiding was mostly occurring in concrete single word concepts (concreteness is equal or above 3), which points towards visual blending being less adequate for such concepts (Section 10.4.2).

Q2: *Is there a difference in the performance of the system between single-word and double-word concepts?*

- running the concept lists, the percentage of concepts with blends is higher in the double-word concepts (77.9%) than in single-word (68%) – (Section 10.3).

- as observed in Table 10.5, for double-word concepts the number of concepts without exported blends was lower (19 vs 27) and the number of concepts with at least one blend without hiding was higher (70 vs 52). These results indicate that visual blending is less useful in the representation of single-word concepts (Section 10.4.2).

- in terms of quality, results showed that there are no statistically significant differences between the representation quality for one word or double-word concepts. However, there are statistically significant differences between single-word and double-word concepts in terms of quality when we consider the number of concepts with hidden emoji (p-value < 0.001887). As we mentioned, there is less hiding for double-word concepts (Section 10.4.2).

- the quantity of different emoji used in the blend production is higher in single-word concepts. This explains why more solutions were exported for single-word concepts and more cases of early satisfaction ($Q \geq 4 \wedge E > 0 \wedge G < 20$) exist for double-word concepts – less emoji thus less variability of solutions and less need for exploration (Section 10.4.2).

Q3: *Is there any connection between the concreteness value of the concepts and the visual concreteness of the emoji gathered by the system to represent it?* We were not able to find any relation between the concepts' concreteness and the visual concreteness of the emoji used in their blends (Section 10.4.2).

11 EMOJINATING: CO-CREATING EMOJI BLENDS

One of the main purposes of *Emojinating* is to be used as a tool for creativity fostering in ideation activities. In such activities, one technique that can be used involves collaboration between different people who share a given goal, leading to what is often referred to as co-creativity. When it comes to computational systems, co-creativity can also be used to achieve unexpected results. However, the relation between user and system is complex. The level of autonomy given to the system directly influences its potential for creative behaviour and degree of contribution to the cooperation with the user.

In the previous chapters, we have shown how *Emojinating* can be used for concept visual representation. However, the developed approaches (described in Chapters 8 and 9) are limited in regards to the collaboration between the user and the system.

The version of the system described in Chapter 9 employed an interactive evolutionary approach that allowed the user to interact with the system and evolve solutions that fit their preferences. Despite this, it could be said that the system is closer to a creativity support tool than to a co-creative system, in the way that the system mostly responds to user requests.

In this chapter, we describe an approach to increase the creative features of the system, with the goal of improving the co-creative relation with the user. The main contributions of this chapter are: (i) the addition of an *automatic evaluator* to the evolutionary process, capable of *self-evaluating* the solutions and adapting to user preferences, (ii) the introduction of *context-adaptation methods*, and (iii) the implementation of a new blend type (*fusion*) and a method for guiding it. We assess the performance of the system by conducting a user study.

This chapter is based on the work presented in the papers by Cunha et al. (2020b) and Cunha, Martins, and Machado (2020a).

11.1 CONTEXT

Over the last few years, several authorss have also addressed the evaluation of creativity in co-creative approaches (Jordanous, 2017; Karimi et al., 2018a). Two aspects are often considered as requirements in a co-creative system: *synchronous collaboration* (Davis et al., 2015) and a *proactive contribution* from both the user and the AI agent (Yannakakis, Liapis, and Alexopoulos, 2014). This means that both agents engage in the interaction and actively contribute to the creative task. Moreover, not only is it required that each agent expresses its own creative ideas but also that it perceives the contributions made by other agents (Karimi et al., 2018a).

Upon the development of creative systems for the visual domain, one of the biggest issues concerns the dependency on human perception – there is no optimal solution as quality depends on the preferences of the user. As already mentioned, one approach that is considered as suitable for such open-ended problems is Interactive Evolutionary Computation (IEC) – an approach used in the previous version of *Emojinating* (Chapter 9).

Nonetheless, when using an Evolutionary Algorithm (EA), one often faces many challenges concerning configuration and parameterisation. For instance, one has to decide how to represent individuals, how to create new individuals and which operators should be used. Additionally, creating a fitness function to evaluate the quality of each individual might not be a trivial task. This is particularly difficult when the fitness depends on the users, their perception and preferences. One possible way to tackle this challenge consists in using a trial-and-error approach, where the practitioner experiments with several configurations and then select one that achieves reasonable good results. The need to remove this trial-and-error process led to the emergence of *adaptive* and *self-adaptive* algorithms. One of the first EAs to introduce this concept was Evolutionary Strategies (ES) (Bäck, Hoffmeister, and Schwefel, 1991). In concrete, ES used a mechanism that adapted the rate with which operators were applied. Over the years many mechanisms have been proposed to adapt components of the EA (Kramer, 2010). Regarding the archive of solutions, there are also mechanisms that can be used to establish an automatic updating process. For example, Vinhas et al. (2016) automatically updates a solution archive based on novelty: individuals are only added to the archive if their quality is above a given threshold and if they are different from the ones already stored in the archive.

Overall, the combination of user interaction and system self-adaptation provides an adequate setup for a co-creative relation between human and computer. Different types of collaboration are accepted in such co-creative systems (e.g. *partnership* or *assistantship*), which vary in terms of complexity of the relationship between human and computer, but also on the level of autonomy given to each of them. Instilling a self-adaptative behaviour to the system may increase its contribution in the co-creative relationship with the user.

11.2 IMPLEMENTATION

The version of *Emojinating* presented in Chapter 9, despite being able to evolve solutions that match the user taste, has a somehow passive behaviour, as the actions of the system are mostly directly triggered by the user. In this section, we describe our approach to enhance the

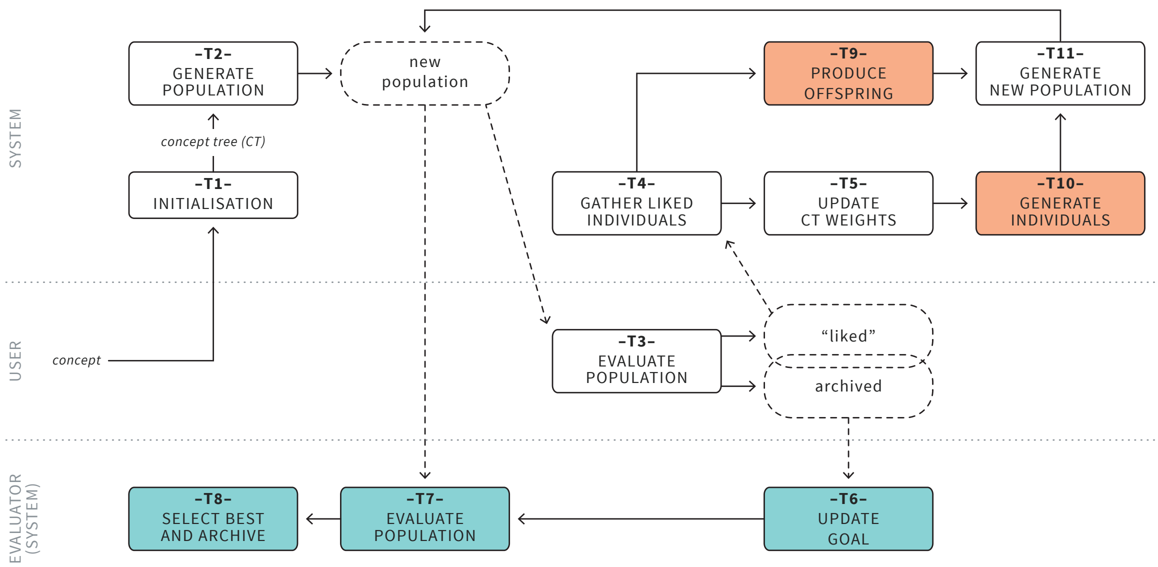


Figure 11.1: Updated version of the Evolutionary framework diagram, showing tasks (T1-11) and objects, e.g. *concept tree* (CT). The figure shows the tasks that correspond to the introduced methods of context-adaptation (tasks T9-10, in orange) and the automatic evaluator (tasks T6-8, in blue). Image best viewed in colour.

creative behaviour of the system, increasing its autonomy and improving the cooperative character of its interaction with the user. In this way, we intend to instil into the system the capability of adequately responding to user actions, thus improving the co-creative relation.

Briefly describing, the previous version of the system could be said to have two agents: an *evaluator* (user) and *solution generator* (system). We now introduce a *second evaluator* (system) that is able to select solutions based on its own idea of quality and storing them in its archive. In addition, we improved the solution generator, increasing its ability to adapt to the context. In Fig. 11.1, the updated diagram of the evolutionary framework is shown, in which T9-10 (in orange) correspond to the production of offspring and generation of individuals, which were modified to include the methods of *context-adaptation*, and T6-8 (in blue) correspond to the added tasks, which are performed by the automatic evaluator. In this chapter, we focus on the changes that were conducted. For a full description of the evolutionary engine, we refer the reader to Section 9.2.

11.2.1 *Representation*

In the previous versions of the system, only *juxtaposition* and *replacement* blend types were implemented. In this version, we improved the blending process by changing the representation used in order to fully include *fusion*, enabling mechanisms of *addition* and *removal*.

The emoji of *Twitter*'s *Twemoji* dataset are composed of layers. We consider the blend as the phenotype of an individual. Each individual is encoded using a genotype of two chromosomes, which codify the combination between the two emoji parents. The emoji used in the blend are stored in the first chromosome. The second chromosome is composed of an undefined number of genes, each codifying an exchange between the two emoji. Each gene corresponds to a set of two numbers (a and b) that refers to the replacement emoji A (A) and to the base emoji B (B), and define how the exchange is conducted. Three different situations may occur: (i) to codify the exchange of layer we use numbers in the 0-1 interval, which correspond to relative position of the layer in the layer array (the number of layers is not the same among emoji); (ii) to codify using the whole emoji instead of the layer, we assign a value of -1; (iii) we use -2 when nothing is to be used of the corresponding emoji. As such, the following cases occur:

- $a = -1 \wedge b = -2$: adds A (*juxtaposition*);
- $a = -1 \wedge b \geq 0$: replaces part of B with A (*replacement*);
- $a \geq 0 \wedge b \geq 0$: replaces part of B with part of A (*fusion*);
- $a \geq 0 \wedge b = -2$: adds part of A (*fusion*);
- $a = -2 \wedge b \geq 0$: removal of part of B (*fusion*).

Figure 11.2: Blend produced for "hot bear" that shows exchanges of the eyes and head shape.

Figure 11.3: Blend produced for "hot bear" that shows an addition of an element (the droplet).

11.2.2 *Implementing Fusion*

This representation approach is different from previous versions in which emoji A (*replacement*) was always used as a whole (no *fusion* was implemented) and, for that reason, only the number referring to emoji B was taken into account (-1 was a *juxtaposition* and ≥ 0 was a *replacement*). The addition of the -2 value made it possible to implement *fusion*, allowing addition, removal and exchange of parts.

One of the main challenges of *fusion* is to know which parts to exchange. Whereas with *replacement* the number of possibilities of blends with only one exchange, despite big, was still manageable (equal to the number of layers of emoji B), with *fusion* this number increases considerably (number of layers in emoji A × number of layers in emoji B). When analysing other examples of visual blending that use fusion we see that the fusion is somehow guided. For example, in the case study described in Chapter 6, the exchange of parts followed conceptual mappings between parts. These mappings were previously produced using a *Mapper* component (Gonçalves, Martins, and Cardoso, 2018) that conducted structural mapping with mental spaces aligned with the input visual representations.

With *Emojinating*, such a method is not appropriate, as no mental spaces are available. Using an unguided fusion would increase the number of nonsense blends, as any combination would be possible.

As such, an alternative method was conceived. We conduct a procedure of finding matches between the parts of the two emoji, based on their position. This is based on the idea that, for some emoji (e.g. *faces*), similar parts are located in a similar position. The *fusion suggestion method* is composed of the following steps:

1. find the nearest part of emoji B for each part of emoji A;
2. solve repetitions by choosing the lowest distance pair;
3. the unmatched parts of emoji A are considered additions;
4. the unmatched parts of emoji B are considered removals.

In each generation, new individuals are generated from scratch using different types of blending. In the case of fusion, after the parent emoji are selected, the *fusion suggestion method* is used to obtain lists of possible *matches*, *additions* and *removals*. The number of exchanges (codified as a *gene*) is randomly set (we used a maximum of four exchanges) and, for each exchange, the system randomly decides between *part exchange* (see Fig. 11.2), *part addition* (see Fig. 11.3) and *part removal*, and randomly selects it from the corresponding list.

When a part is exchanged, the replacement part assumes the size and position of the replaced part. When an addition occurs, the added part is placed in its original position and no resizing occurs. The fusion suggestion method is only used in the generation and not in the mutation to reduce nonsense blends but still allow more variation.

We are aware that this does not work for every emoji but fusion is also not suitable for every combination – it usually works better with emoji that are visually and structurally similar. Despite not being a perfect mechanism, it solves some problems in identifying the parts. Other approaches may be explored in the future.

11.2.3 *Variation Operators*

The system presents the user with a population of a previously defined number of individuals (we used a population with 20 individuals), i.e. blends, and in each generation the user selects the ones to go through a process of producing offspring (T9 in Fig. 11.1). Two different operators exist: *crossover* and *mutation*. The produced offspring individuals from both operators are added to an *offspring pool*. A maximum percentage of the new population (50%) is reserved for the offspring, which are randomly selected from the pool. The remaining percentage corresponds to individuals generated from scratch.

11.2.3.1 *Crossover Operator*

A *crossover* occurs when the user selects at least two individuals. Initially, the system only conducted crossover with individuals that shared at least one emoji. Afterwards, we realised this approach severely reduced the possible offspring. As such, we decided that blends with no shared emoji could also be combined.

In order to conduct the crossover, groups of two are randomly made with the emoji selected by the user. Two types of crossover can occur: if the number of exchange genes (second chromosome) is equal to one, one of the emoji of each parent individual is exchanged with the other individual; if the number of exchange genes in both emoji is above one, it conducts a gene crossover. A *gene crossover* consists in exchanging genes between individuals, using a *one-point crossover*. The resulting offspring individuals are added to the offspring pool.

11.2.3.2 *Mutation Operators*

In the previous version of the system (described in Chapter 9) only three types of mutation existed (see Section 9.2.1.6): *replacement emoji*, *replaced layer* and *blend type*. With the new representation, the types of mutation increased to the following seven:

- *replacement emoji:* the emoji used as replacement is changed;
- *base emoji:* the emoji used as base is changed;
- *juxtaposition:* the blend type changes to juxtaposition;
- *replaced layer*: the replaced layer is changed;
- *replacement whole to layer:* the replacement part changes from the whole emoji (a = -1) to a layer ($a \geq 0$);
- *replacement layer to layer:* the layer used as replacement part is changed;
- *replacement layer to whole:* the replacement part is changed from a layer ($a \geq 0$) to the whole emoji (a = -1).

11.2.4 *Adaptation*

Two types of adaptation can be said to exist: (i) to the user and (ii) to situations within the system (context). In the evolutionary approach, we introduced mechanisms that are aligned with the former, making the system adapt to the user preferences (see Section 9.2.1.3). One of the goals of the approach described in this chapter is to focus on the latter, allowing the system to adapt to the population at the moment, as different stages in the run may require different behaviour from the system. Two different means of context-adaptation were implemented: *adaptive blending process* (individual generation) and *adaptive variation operators* (mutation).

11.2.4.1 *Adaptive Blending*

The adaptive blending process consists in changing the likelihood of a given type of blend occurring, according to the state of the population. This is used in the generation of new individuals from scratch (T10 in Fig. 11.1). The types of blend have different variation potential (juxtaposition has the lowest potential and fusion has the highest). Due to this, our approach is that blend types with higher variation potential should occur more frequently when there are fewer different emoji used in the blends of the population. As such, we assign the probability of each blend type based on the number of different emoji (N_E):

- for N_E ≥ 20 (higher), JUX $= 10\%$ and FUS $= 20\%$;
- for N_E ≤ 8 (lower), JUX $= 2\%$ and FUS $= 50\%$;
- for $8 <$ N_E < 20, the probabilities are calculated with the following expression:

$$\text{LOWER\_VAL} + (\text{UPPER\_VAL} - \text{LOWER\_VAL}) \times \frac{(\text{N\_E} - 8)}{12}, \tag{11.1}$$

where LOWER_VAL and UPPER_VAL are the probability values of the blend type used in the lowest and highest bounds of N_E (e.g. in JUX 2% and 10%, respectively). The value for replacement is calculated using the following equation:

$$\text{REP} = 100 - \text{JUX} - \text{FUS}. \tag{11.2}$$

The values used were empirically obtained through experimentation and adjustment.

11.2.4.2 *Adaptive Mutation*

Regarding mutation adaptation, our initial approach was similar to the adaptive blending process: we tried to assign the same value to each operator and change it according to the state of the population. Later we concluded that due to the characteristics of the problem, this approach would not lead to good results.

We realised that each type of blend has its own particularities and, therefore, has different mutation requirements. For example, in *juxtaposition* mutating the replaced emoji is simple as the whole emoji is used, whereas in *fusion* it is more complex as the layer-based exchanges are relative to the array of layers of each emoji, which varies in number of elements – mutating the replaced emoji in fusion would result in something entirely different, causing a huge change. As such, mutation adaptation consists in changing the occurrence probability of each

Table 11.1: Probability of each mutation type based on the type of blend (JUX, REP OR FUS) of the individual being mutated. RP stands for Replacement Part.*Implemented but not used.

mutation type	JUX	REP	FUS
replacement emoji is changed	40	30	5
base emoji is changed	40	10	0
*blend type changes to juxtaposition**	0	0	0
replaced layer is changed	20	35	25
RP changes from whole to a layer	10	5	0
RP is changed by selecting a new layer	0	0	15
RP changes from a layer to whole	0	0	5

mutation operator according to the blend type of the individual being mutated. We established values for each mutation, depending on the type of blend of the parent (see Table 11.1). The emoji mutations are independent of the rest. If *juxtaposition* occurs, none of the rest occurs. If no *juxtaposition* occurs, any of the other mutation types can occur. These changes were conducted in the task of producing offspring (T9 in Fig. 11.1).

11.2.5 *Self-evaluation and Selection*

In order to give some autonomy to the system, we decided to bring another agent to the evolutionary process. This agent is an automatic evaluator that has two possible actions: evaluate individuals according to its preferences and store individuals in its own archive. The user can see the archive and is able to retrieve individuals from it but only the automatic evaluator is able to add individuals.

11.2.5.1 *Quality Assessment: Criteria*

Defining criteria for quality assessment of blends is not an easy task. First because quality is dependent on visual attributes but also on conceptual ones (e.g. does the user perceive the concept?). Moreover, as they depend on user understanding and perception it makes this an open-ended evolution problem.

We chose to focus on the first type of criteria (visual attributes). We considered two aspects that are related to the quality of an icon: complexity (the simpler the better) and legibility (should be perceivable in smaller sizes). Also, given that we are conducting visual blending, we need to also consider the degree of change in comparison to the parents. With this in mind, we defined the following criteria:

1. overall complexity (*oCom*):

$$oCom = \frac{1}{\#\text{BLEND\_LAYERS}}. \tag{11.3}$$

2. area exchanged (aEx):

$$aEx = \sum_{i=1}^{\#LE(b)} a(l_i), \tag{11.4}$$

where $\#LE$ is the number of layers exchanged (added+removed) in the blend (b) and function a calculates the area of a layer l.

3. relation between added area and added layers (rAL):

$$rAL = \frac{\sum_{i=1}^{\#LA(b)} a(l_i)}{\#\text{ADDED\_LAYERS}}, \tag{11.5}$$

where $\#LA$ the number of added layers and function a calculates the area of a layer l.

4. difference in complexity ($cDif$):

$$cDif = \#\text{ADDED\_LAYERS} - \#\text{REMOVED\_LAYERS}. \tag{11.6}$$

11.2.5.2 *Quality Assessment: Fitness Calculation*

The goal of the automatic evaluator is to be able to assess solutions based on its own idea of quality. In this sense, there are two options: having an evaluator that tries to get similar solutions to the user, in order to present good alternative solutions; or get solutions that are distinct from what the user is selecting. In our implementation, we chose to focus on the first approach.

The system's idea of good solutions is therefore dependent on user choices. This is achieved by making the system analyse the blends in the user archive – which are assumed as being good – and afterwards change its idea of a good solution to match these user-selected blends. In the beginning, the system starts with default values (all equal to 1). As soon as the user stores individuals in the archive, the system evaluates them and changes its fitness goal, based on their characteristics. To obtain the goal, the system calculates the average of each criterion for the individuals stored in the user archive, which results in an average blend profile (T6 in Fig. 11.1). This profile is then set as the new goal and used for selecting individuals that the system finds interesting. As such, the system goal changes over time, according to user preferences. In order to calculate the fitness of an individual, the system uses a *Euclidean distance* between the individual and the average profile:

$$d = \sqrt{\sum_{i=1}^{4} (c_i - \mu_{c_i})^2}, \tag{11.7}$$

where c_i and μ_{c_i} denote the criterion of the individual being evaluated and the average value of the criterion, respectively. With this function, the system assesses how far it is from the goal. The average profile is updated at the end of each generation.

11.2.5.3 *Individual Selection*

As already mentioned, the automatic evaluator has its own archive. At the end of each generation, and after calculating the new goal (T6 in Fig. 11.1), the system performs an analysis of the population to check for good individuals (T7 in Fig. 11.1). The evaluator's archive capacity was set to 5 to avoid storing too many individuals. Also to avoid collecting too many individuals, the evaluator only stores one blend per emoji combination. This way, the system tries to improve the fitness of individuals for each emoji combination. In each generation, the evaluator selects the best blend in the population (T8 in Fig. 11.1), checks whether it already has a blend for the emoji combination and proceeds as follows:

1. If there is already a blend for the emoji combination:
   a) If the fitness of the stored one is lower than the current population best, it replaces the individual in the archive;
   b) If the stored individual has higher fitness, the evaluator discards the selected individual and proceeds to the next generation without storing any individual.
2. If it does not have any blend for the emoji combination:
   a) If the archive has free space, the evaluator stores the blend;
   b) If the archive is full, the evaluator checks whether the selected blend (best blend in the population) has higher fitness than the worst individual in its archive. If so the evaluator replaces the stored blend with the selected one; otherwise, the selected blend is discarded.

In each generation, the evaluator discards individuals from its archive if the difference of their fitness to the best individual in the population is greater than two (empirically obtained), which often happens when the fitness goal drastically changes.

11.3 EXPERIMENTATION

In order to assess the performance of the new version of the system (referred to as co-creative), we conducted a user study.[1] The study was conducted with eight users, who also participated in the experiment described in Section 10.4. The two studies were separated by a two-month period. Therefore, we consider the present study as a follow-up to the first one, allowing us to compare the two versions of the system (*evolutionary* vs *co-creative*). The previous results (Section 10.4) suggested that our system is not very suitable for the representation

1 This study has been partially described in the publications by Cunha et al. (2019b) (section User Study #1) and Cunha et al. (2020b) (section Experiment #2).

of single-word concepts, especially concrete ones (e.g. *dog*). For this reason, we chose to focus on double-word concepts.

In this section, we will compare the results of the two studies, using stage #1 (ST1) to refer to the results obtained in the study described in Section 10.4 and stage #2 (ST2) to refer to the ones from the study described in the present section.

11.3.1 *Experiment Setup*

For the results of ST1, we only use data of participants who were in both studies (eight participants).

In ST2, the participants used the co-creative approach described in the previous section and tested the same double-word concepts used in ST1, each testing a total of five concepts. As described in Section 10.4, the concepts were randomly selected from a list built by crossing a noun-noun compound dataset (Fares, 2016) with a concreteness ratings dataset (Brysbaert, Warriner, and Kuperman, 2014), which was divided into groups based on semantic concreteness. Each participant tested at least one concept from each concreteness group – *very low concreteness* (1-2), *low concreteness* (2-3), *high concreteness* (3-4) and *very high concreteness* (4-5).

The participants were asked to use the system to evolve blends that, in their opinion, represented the concept. In stage 1, for each concept the users were requested to conduct three tasks (these tasks were already listed in Section 10.4 but we repeat them here for clarity):

T1 Classify how well the system represented the concept, using a scale from 1 (very bad) to 5 (very good);

T2 Classify the surprise degree of the results, using a scale from 1 (very low) to 5 (very high);

T3 Export the solutions which they considered the best, from the ones which they considered as concept-representative. In case no solution represents the concept, none should be exported.

In stage 2, the same tasks were conducted plus an additional one:

T4 Classify the capability of the system to adapt to user actions, using a scale from 1 (very low) to 5 (very high).

We are aware that there is a risk of obtaining biased results as the users had used the same concepts with previous version of the system. This may have affected the results of some of the tasks conducted (e.g. T2 surprise assessment).

The objective of this study is two-fold: (i) compare the two versions of the system, and (ii) compare the different types of blend and assess the impact of adding fusion.

Table 11.2: Results of the two stages (ST1 and ST2) for T1 (quality), expressed in number of concepts and divided into groups based on number of retrieved emoji (*small*, *medium* and *large*).

	quality					
	1		>1 and ≤3		≥4	
	ST1	ST2	ST1	ST2	ST1	ST2
small	1	1	5	3	4	6
medium	6	4	3	3	9	11
large	1	2	3	0	8	10
	8	7	11	6	21	27

Table 11.3: Results of the two stages (ST1 and ST2) for T3 (blends exported), expressed in number of concepts and divided into groups based on number of retrieved emoji (*small*, *medium* and *large*).

	exported					
	0		1		>1	
	ST1	ST2	ST1	ST2	ST1	ST2
small	1	2	8	7	1	1
medium	8	7	6	8	4	3
large	1	2	6	6	5	4
	10	11	20	21	10	8

We also analysed the total number of emoji retrieved for each concept and divided the concepts into three groups: *small* (emoji number ≤5), *medium* (>5 and ≤15) and *large* (≥25). This was done to assess if to what extent the performance of the system is dependent on the quantity of available emoji. Each of the participants had at least one concept from each group.

11.3.2 *Results and discussion*

The results show that the new approach (tested in ST2) led to an improvement in performance – the number of concepts with quality lower than good ($= 1$ and > 1 and ≤ 3) decreased and of quality equal or higher than good (≥ 4) increased from 21 to 27 (Table 11.2). No major differences are observed in relation to the different groups (emoji quantity). When comparing the two stages (ST1 and ST2) in terms of concept representation (Table 11.3), the difference is small – ST1 had a total of 30 represented concepts (with exported blend) and ST2 had 29 (3 lost and 2 gained).

In terms of exported blends, the participants exported one blend in the majority of the concepts (Table 11.3). In ST2, from the 40 concepts, we obtained the following results: no solution was exported in 11; in 21

Table 11.4: Results of the two stages (ST1 and ST2) for T2 (surprise), expressed in number of concepts and divided into groups based on number of retrieved emoji (*small*, *medium* and *large*).

	surprise					
	1		>1 and ≤3		≥4	
	ST1	ST2	ST1	ST2	ST1	ST2
small	1	2	5	4	4	4
medium	2	0	8	9	8	9
large	0	0	8	4	4	8
	3	2	21	17	16	21

Table 11.5: Results of ST2 for T4 (adaptation), expressed in number of concepts and divided into groups based on number of retrieved emoji (*small*, *medium* and *large*).

	adaptation		
	1	>1 and ≤3	≥4
small	2	4	4
medium	0	7	11
large	0	5	7
	2	16	22

only one solution was exported; and in eight more than one solution was exported (see examples of exported blends in Figs. 11.4 and 12.1).

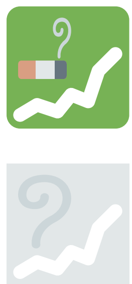


Figure 11.4: Blends produced for "cigarette market".

The fact that the process of visual blending was able to lead to good solutions for the majority of the concepts seems to indicate that it is a useful method for concept representation. Moreover, the results also show that the system is able to present the user with more than one good solution.

We also analysed three combinations of quality (Q), exported (E), and generation (G) that correspond to "early quit without results" ($E{=}0 \wedge G{<}20$), "early quit with poor results" ($Q{\leq}3 \wedge E{>}0 \wedge G{<}20$) and "early satisfaction" ($Q{\geq}4 \wedge E{>}0 \wedge G{<}20$). The total number of early satisfaction was maintained, increasing in the ones with few emoji – indicates that the system is able to produce solutions of good quality that the previous version could not – and decreasing in the medium and large groups – indicates that the user was still exploring the space or that the user was not able to obtain what he wanted. The number of early quitting with poor results decreased (maybe due to more variety). However, the number of early quit without results increased by 2 in the small emoji quantity group. A possible interpretation for this latest result is that the participants may have remembered that the system did not perform well for that concept in ST1, which caused a quicker quit.

Table 11.6: Results of the two stages (ST1 and ST2) in number of occurrences (exported blends) of each type of blend – *juxtaposition* (JUX), *replacement* (REP) and *fusion* (FUS) – divided into groups based on the number of retrieved emoji (*small*, *medium* and *large*). The "?" column refers to cases in which it was not possible to identify the type of blend and "hidden" to cases in which one of the emoji was hidden.

	ST1					ST2				
	JUX	*REP*	*FUS*	*?*	*Hidden*	*JUX*	*REP*	*FUS*	*?*	*Hidden*
small	6	4	–	0	2	2	3	1	3	3
medium	10	8	–	0	2	8	6	0	1	2
large	9	9	–	0	1	4	9	2	0	2
	25	21	–	0	5	14	18	3	4	7

As it was not the first time using the system with those concepts, we anticipated that the participants would not be as surprised as in ST1. However, the results show the opposite: there was an increase in the surprise values (Table 11.4). The number of concepts with surprise ≥ 4 increased from 16 to 21, which means that the system was still able to generate surprising results.

In terms of adaptation (Table 11.5), in the majority of concepts (22) it was reported as equal or above high ($\geq$4). There were only six cases of adaptation ≤ 2. Only in two of those cases did the user go beyond the tenth generation. From the remaining four, only one of them was above the fifth generation (G= 7). Moreover, the user was only able to find a satisfying solution (i.e. the participant exported one solution) in one of these four cases. In these cases, the participants stopped the run very early (e.g. generation four) mostly due to lack of emoji, and this affected their perception of the system's capability of adaptation.

To further investigate the suitability of visual blending for concept representation, we analysed the blends exported by the participants in terms of blend type (some concepts had more than one exported). In total, ST1 had 46 exported blends and ST2 had 39 (Table 11.6).

Figure 11.5: Blend produced for "airline bureaucracy".

Considering that *fusion* is only used in ST2, the results show that *juxtaposition* and *replacement* are used in the majority of the exported blends (Table 11.6). In addition, in some cases of ST2, it was not possible to ascertain the type of blend, as one of the emoji was hidden. Another emoji hidden situation occurred in a *fusion* blend, in which the replacement emoji was not perceivable (Fig. 11.5). We identified the cases in which one of the emoji was hidden in the blend (Table 11.6). These results seem to indicate that *fusion* is not very useful and may only add unnecessary variability to the results.

11.4 GENERAL ANALYSIS

In general, the system is able to learn from the user behaviour, which is observed in the storing of similar blends in its archive (e.g. if the user selects blends with a large exchanged area, the system tends to replace the blends in its archive to match the user preference). Interestingly, some participants reported that the system was placing good and previously unseen solutions in its archive, and some even stored blends gathered from the system's archive, which shows that the system was able to adapt to user preferences. This points out that the system archive is useful to highlight blends that the user may have missed as the system only selects blends that were previously shown to the user. It also indicates that the number of individuals shown to the user (equal to population size) may be too large to be efficiently analysed, as the users seem to miss some blends that they consider good. On the other hand, some participants mentioned that they did not look much into the system's archive due to its location on the page and to the opacity used in its blends.

Concerning weariness, the obtained values were lower in ST2 (median = 1, mode = 1) comparatively to ST1 (median = 2, mode = 2). However, due to differences in the number of tested concepts and consequent duration of the test, ST1 and ST2 cannot be realistically compared in terms of weariness. Despite this, two participants who conducted the tests with a similar duration in both stages (30 vs 29 and 25 vs 23), reported weariness of 2 in ST1 and weariness of 1 in ST2. This suggests that the version of the system used in ST2 may cause less weariness. Moreover, participants reported that the fact that the system was capable of generating more variation led to more exploration and, consequently, to less monotony and fatigue, which in ST1 was often caused by over-similarity among blends.

Regarding the conduction of the user survey, an aspect should be mentioned. As already stated, the concepts were retrieved from a noun-noun dataset. The concepts, however, are not the most suitable to test the system with as some of them tend to be too specific (e.g. "summit agreement", "interest abatement", etc.) and, in a normal situation, the system would most likely not be used with such concepts. This should be taken into account when interpreting the results.

11.4.1 *Setup and Methods*

The values used in the co-creative approach were empirically obtained through experimentation and adjustments. However, due to the high number of parameters we consider that further tuning is required. One example is the probability of *fusion*, which depends on the number of different emoji in the population. Its probability of occurrence was set to a high value (50%) for low emoji numbers, as it is the type of blend that leads to the highest variety of results – theoretically, this would be suitable in situations in which few emoji exist. However, this does not work when put into practice as it makes it harder to identify both parent emoji (e.g. "airline bureaucracy" in Fig. 11.5), which worsens the user perception from the first generation. A possible solution may be to also consider the number of the current generation.

Figure 11.6: Blend produced for "hot bear" that shows issues with fusion suggestion.

Another issue has to do with the methods used in the system, for example, the *fusion suggestion*. We consider that this method is not fail-proof (see Fig. 11.6 in which the smiley's mouth is paired with the bear's nose) but was one solution that is suitable to solve some of the issues. Nonetheless, there are some aspects that need to be taken into account in future developments. For example, it should have a distance threshold – at the moment it only leaves out parts if the emoji do not have the same number of layers. Moreover, in addition to *position*, it should also consider *size* – two different sized layers may have a similar position (measured in the left corner of the object) and still not be the most suitable for exchange. Another approach to improve the system is to assess which type of blend is the most suitable for the emoji in question (e.g. *fusion* is suitable for blending faces but for blending an animal and a location, *juxtaposition* should work better). This is aligned with some of the conclusions from the study with visual blends described in Chapter 5, in which we identified certain types of transformations that are commonly used in the blends involving specific emoji.

11.4.2 *Interface*

In comparison to the previous version, the interface was improved by reducing the sizes of most elements (Fig. 11.7). This was done to enable the creation of the automatic evaluator archive. Despite this, with the user survey, we identified some issues. First, the position of the archive in the lower part of the page makes it less noticeable. This is further intensified due to the choice of adding some transparency to the blends in the evaluator archive. The reduced presence of the archive fails with the goal of showing blends to the user – some participants stated that they were not paying much attention to the evaluator archive.

11.4.3 *Evaluator Archive*

On the other hand, some participants claimed that the evaluator was able to get good blends that they had not seen before. This highlights two things: the evaluator is able to select individuals that match user preference; the user had missed blends that had previously appeared in the population (maybe due to population size) – the evaluator only selects blends from the population. As such, the evaluator archive is useful to show good solutions that may have been missed.

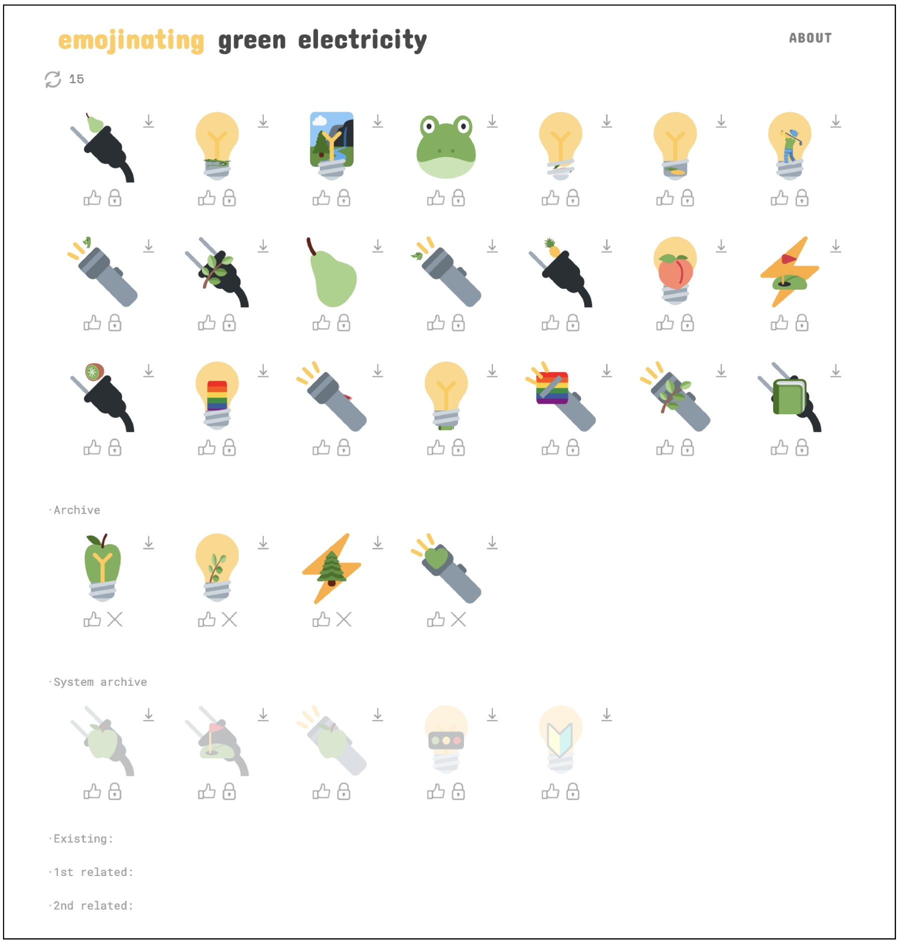


Figure 11.7: Interface of *Emojinating* version 3.

Analysing how the evaluator stores the individuals, we believe that there are some improvements that need to be made. First, in some cases, due to lack of emoji, the evaluator only stored one individual – the best individual for the emoji combination. A possible improvement could consist in allowing the evaluator to add different blends from the same emoji parents until the maximum size is reached. Moreover, the system stores blends that are equal to the ones selected by the user. Ideally, the evaluator should avoid blends that are exactly the same as any stored by the user, as they are not useful in any way.

Concerning the evaluator behaviour, we explored a strategy focused on searching for blends similar to the ones selected by the user. Another possible approach is to use the evaluator to explore different areas of the search space, stimulating a divergent behaviour, which may increase the variety of good results.

11.4.4 *Fitness*

Regarding fitness, there are also some issues that need to be addressed. First, using an average of individuals' properties as a goal does not always work well – an individual located between two good individuals may not be a good individual. Even more, as we are dealing with symbols, in which the perception of the concept is more important than some of the considered visual features (e.g. *area changed*). In addition to visual features, other aspects can be considered – e.g. the emoji being used (aligned with the ones from the blends selected by the user or more closely related to the concept being represented), type of blend (to match user preferences), or even similarity measured using a *pixel-based* approach.

11.4.5 *Role of the Evaluator*

In the current approach, the evaluator is left with the blends that the user did not select. Despite being useful to identify good solutions that the user missed, its role can still be considered as passive. A possible future direction is to allow the evaluator to select blends to be reproduced, generating offspring from its own stored individuals. This can further enhance the system's creative behaviour.

11.5 SUMMARY

In this chapter, we presented an approach to improve the relation between the *Emojinating* system and its users. The previous version was limited in terms of participation of the system in the co-creative exchange with the user. Our goal was to increase the creative behaviour of the system by introducing two novel functionalities: *self-evaluation* and *context-adaptation*. With this improved version, a new agent is introduced in the evolutionary process: an *automatic evaluator*, capable of selecting individuals according to its idea of good quality.

We conducted a user study, which we consider as a follow up of the study described in Section 10.4 – a group of eight people participated in both studies. This way, we were able to compare the new version of the system with the previous one. The results show improvement in the quality of solutions and in the capability of stimulating the user.

12 USEFULNESS, PERCEPTION AND OVERALL ANALYSIS OF EMOJINATING

In the previous chapters, we have presented the three versions of *Emojinating*: deterministic (Chapter 8), evolutionary (Chapter 9) and co-creative (Chapter 11). In addition to the multiple studies that were conducted to assess the performance of the different versions, in Chapter 10, we described a broader study in which we evaluate the quality of visual representation of single-word and double-word concepts.

In this chapter, we start by making a general summary of the iterative development of *Emojinating*. Then, we describe an Experiment divided into two parts, each with a specific goal: (i) comparing creative production by the user alone with results obtained with the system and assessing the perception of quality by the user, simulating a real-world situation (part #1); and (ii) assessing how blends are perceived by people not involved in their production (part #2). We conclude the chapter with an overall discussion on *Emojinating*'s performance and its potential uses.

This chapter is based on the work presented in the papers by Cunha et al. (2020b) and Cunha et al. (2019b).

12.1 RECAP OF EMOJINATING DEVELOPMENT

The iterative development of *Emojinating* was described in the previous chapters. Despite using different approaches, the main goal of the system is always the same: using visual blending of emoji to represent concepts. In general, the system receives a concept from the user that is mapped to two emoji (e.g. emoji A and emoji B), which are then combined through a process of visual blending – emoji B is considered as the *base* for the blending and emoji A as the *replacement*. Three different types of blending were used: *juxtaposition*, *replacement* and *fusion*.

The second version of the system included an interactive evolutionary engine (Chapter 9), which allows the production of solutions that match the user preference. The third version introduced improvements to the collaboration between user and system (Chapter 11).

In these three versions, we have used different approaches for choosing the type of blend and replaced parts: in the *deterministic* system (Chapter 8) one blend was generated for *juxtaposition* and for *replacement* the system generated one blend for each layer replacement possibility, per emoji combination; in the *evolutionary* system (Chapter 9), for the generation of new blends, we used a rate of 20% for *juxtaposition* and 80% for *replacement* (the replaced layer was randomly chosen); and

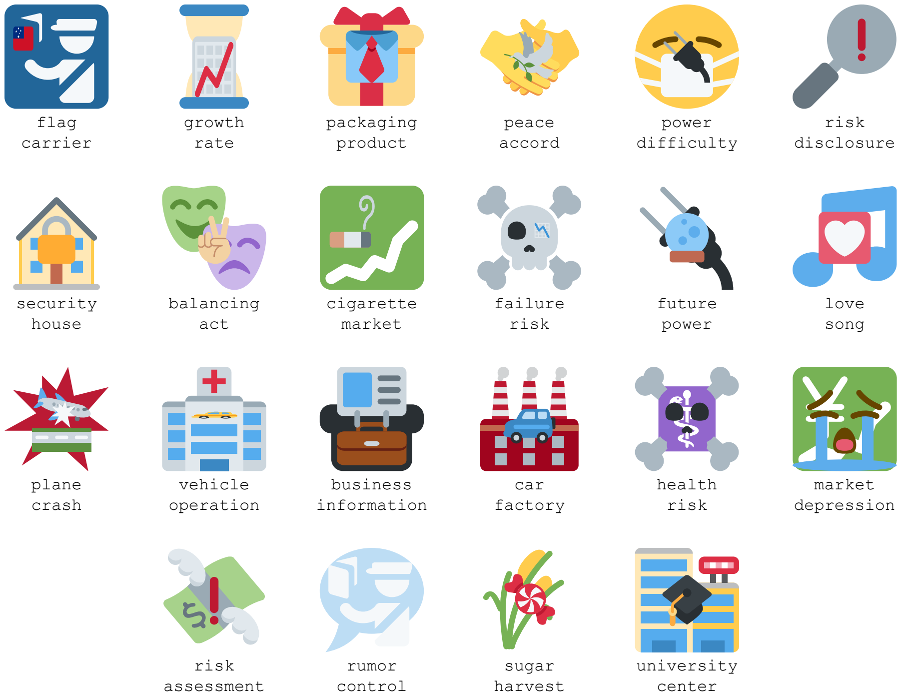


Figure 12.1: Blends obtained in the study described in Section 11.3.

in the *co-creative* system (Chapter 11), *context-adaption* was used (changing the blend type probabilities depending on the population) and the choice of the replaced layer is made based on perceptual features and user preference (e.g. if the user prefers small replaced parts, the system will tend to produce blends that match this preference). The goal of this iterative development was to improve the way that the system is used to produce visual representations of concepts. However, one may question whether the system can actually be useful to users.

12.2 ASSESSING USEFULNESS

The capability of finding a solution that visually represents a given concept does not necessarily present evidence of the usefulness of the system itself. Moreover, the perception of a designer does not always match how their design is perceived – i.e. even though a user might consider that a blend has high quality, this does not mean it will be well interpreted by someone who does not know the concept behind it. In this section, we describe an experiment[1] in which we focus on the assessment of the potential usefulness of the system and also the perception of blends by users.

We used blends exported by the eight participants of the study described in Section 11.3 (Fig. 12.1). From all blends exported, we selected

1 This experiment has been (partially) described in the publications by Cunha et al. (2019b) (section User Study #2) and Cunha et al. (2020b) (section Experiment #3).

only one per concept, using the ones identified as the best when more than one had been exported. Then, we excluded the ones in which one of the emoji was being hidden, as these could not be considered as proper visual blends. This resulted in a set of 22 concepts and corresponding visual representations (Fig. 12.1). The set was divided into three groups, balanced in terms of *semantic concreteness*.

Three sessions were conducted, resulting in three user groups: group 1 with 15 participants, group 2 with 19 and group 3 with 22 (Table 12.1). This difference in participant number was due to participant availability. In total, 56 users with ages between 19 and 27 (*average* = 20.4 and *standard_deviation* = 1.6) participated in the experiment, all with a background in graphic design.

12.2.1 *Part #1 Usefulness*

In the first part of this experiment, we focused on assessing the potential usefulness of the system.

12.2.1.1 *Experiment Setup*

Each group of concepts was given to a group of participants – each concept was tested with a minimum of 15 users.

Each participant received a list of concepts and had to complete a survey for each concept. The survey was divided into two stages and was composed of five tasks. First, each participant was asked to conduct four tasks for each concept:

T1 Do you understand the concept?

T2 Draw the concept.

T3 Describe the drawing in a few words.

T4 How well does the drawing represent the concept?

Tasks T1 and T4 required the participant to use a scale from 1 (not at all) to 5 (perfectly). In case the participant did not understand the concept (T1), the remaining tasks were to be ignored. The participants were told to use a quick drawing style, similar to the one used in games such as *Pictionary* (see user drawings in Fig. 12.2). After conducting the four tasks for every concept, the generated blends of each concept were shown and the participant was asked to answer the following question for each concept, using the previously described 1-5 scale:

T5 How well does the blend represent the concept?

12.2.1.2 *Results*

The study resulted in a total of 414 concept tests – group 1 was composed of seven concepts and was tested with 15 participants; group

Table 12.1: Results for each concept – average semantic concreteness (C); compositionality score (CP) obtained from the work by Roberts and Egg (2018); number of tests conducted (T); mode (*mo*) and median ($\tilde{x}$) for the tasks T1 (concept understanding), T4 (drawing quality) and T5 (blend quality); number of valid tests (V); number of tests analysed (A); concepts with good understanding in the majority of the valid tests are marked with · in column U; percentage of A in which the blend was worse (B $<$ D) and better (B $>$ D) than the drawing (includes absence of drawing). We use bold to identify the winning approach and underline to identify concepts with high understanding rate in which the blend was better than the drawing. The values in C an CP were rounded to two decimal places for better presentation.

				T1		T4		T5							
	C	CP	T	*mo*	$\tilde{x}$	*mo*	$\tilde{x}$	*mo*	$\tilde{x}$	V	A	U	B<D	B=D	B>D
flag carrier	4.50	0.35	15	1	2	1	2	4	3	10	1		0.0	0.0	**100.0**
growth rate	2.70	0.63	15	4	4	4	4	5	5	14	10	·	20.0	20.0	**60.0**
packaging product	4.31	-	15	5	5	3	3	4	3.5	13	9	·	33.3	22.2	**44.4**
peace accord	1.60	0.47	15	4	4	3	3	5	5	14	10	·	10.0	20.0	**70.0**
power difficulty	1.97	-	15	1	2	1	1.5	3	2.5	11	3		**100.0**	0.0	0.0
risk disclosure	1.91	-	15	1	1	1	1	3	3	10	1		**100.0**	0.0	0.0
security house	3.91	-	15	3	3	3	3	5	4	11	6	·	33.3	16.7	**50.0**
balancing act	2.82	0.38	19	5	3	3	3	1	1.5	14	7		**71.4**	14.3	14.3
cigarette market	4.79	-	19	5	4	3	3	4	3	17	9	·	33.3	22.2	**44.4**
failure risk	1.86	-	19	5	3.5	2	2	1	3	15	6		**50.0**	33.3	16.7
future power	1.95	-	19	5	3	1	3	4	3	13	6		16.7	33.3	**50.0**
love song	3.27	0.58	19	5	5	5	4	5	5	16	15	·	6.7	**53.3**	40.0
plane crash	4.36	0.44	19	5	5	5	4	5	5	13	12	·	25.0	25.0	**50.0**
vehicle operation	4.04	-	19	3	3	4	3	3	3	14	6		**50.0**	33.3	16.7
business information	3.08	0.47	22	3	3	1	2	4	4	18	2		0.0	0.0	**100.0**
car factory	4.83	0.48	22	5	5	2	2.5	5	5	20	16	·	6.3	18.8	**75.0**
health risk	1.96	0.49	22	4	4	2	2	5	4	20	18	·	11.1	11.1	**77.8**
market depression	3.55	-	22	1	2.5	2	2	1	2	20	5		**60.0**	20.0	20.0
risk assessment	1.87	0.57	22	1	1	3	3	3	3	18	2		50.0	0.0	50.0
rumor control	2.04	-	22	1	3	3	3	3	3	21	3		**100.0**	0.0	0.0
sugar harvest	4.44	-	22	1	1	1	2.5	4	4	17	3		33.3	33.3	33.3
university center	4.25	0.59	22	1	3	2	2	5	4	19	8		0.0	0.0	**100.0**

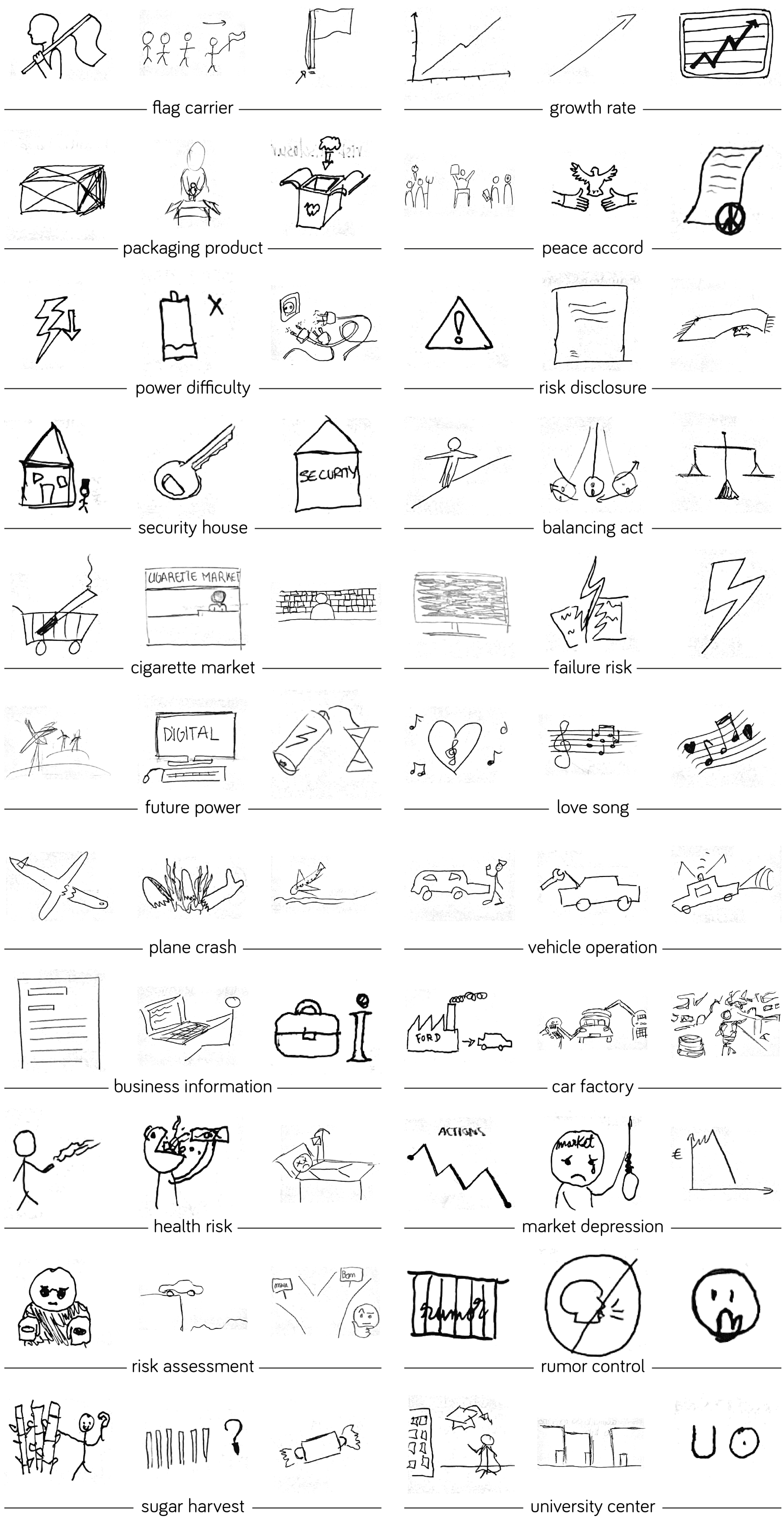


Figure 12.2: Examples of user drawings of each concept

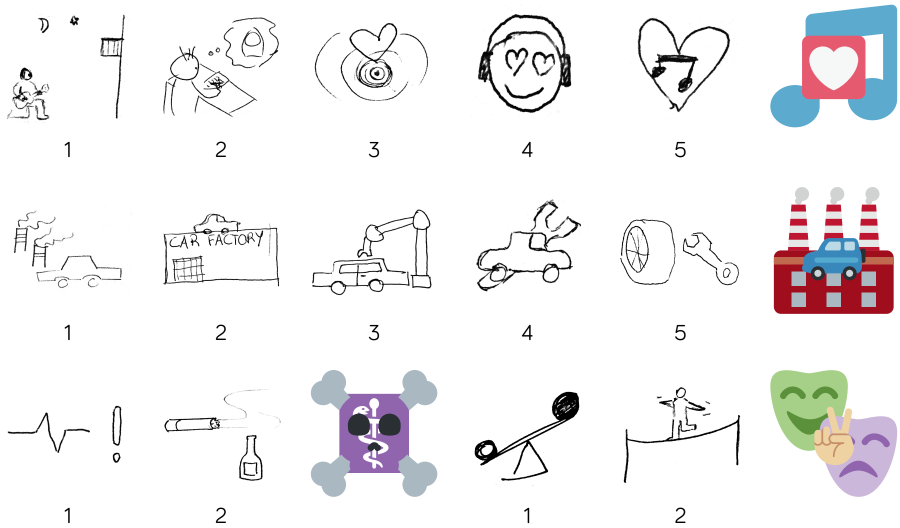


Figure 12.3: User drawings obtained for *love song* (top row), *car factory* (middle row), *health risk* (bottom row, left side) and *balancing act* (bottom row, right side).

2 had seven concepts and was tested with 19; and group 3 had eight concepts and was tested with 22. From the total of tests, 76 tests had to be excluded from the study due to invalid answering: in three tests no answer was given to any of the tasks; in 24 no answer was given regarding the familiarity with the concept (T1); in 40 the quality of the blend was not evaluated (T5); and in nine a visual representation was drawn but not evaluated (T4).

In addition to these validity exclusions, for our analysis we only considered tests in which the participant reported to know the concept well or perfectly (T1 $\geq$4). We are aware that this procedure reduced the number of answers considerably but it would be illogical to analyse tests in which the concept was not known to the participant – such would invalidate the results. It is important to notice the difference between valid tests (V) and valid tests in which the user understood the concept (A) – see Table 12.1. This may be due to two factors: concept complexity and participant language difficulties (the participants were not native English speakers).

The results of the comparison between blend and drawing shown in Table 12.1 only consider the valid tests in which the user understood the concept (A). The cases in which no drawing was made are included in the value of better blend (B > D). Comparing the results from T4 (drawing) and T5 (blend) allows us to assess whether the system can be useful.

In eight out of the nine concepts with high understanding rates – i.e. good understanding in the majority of the valid tests (marked in column U in Table 12.1) – the blend was considered better than the drawing by the majority of the participants (underlined in Table 12.1) and this majority was even absolute in four from these concepts (*growth* *rate*, *peace accord*, *car factory* and *health risk*). From the remaining 13 concepts, in four the blend was considered better by the majority of the participants, and in two the results obtained by the blend and the drawing were equal. In contrast, the drawing was only better than the blend for the absolute majority of A in five concepts, four from which had a very low understanding of the concept (less than 27% had T1≥4). Moreover, in two cases, the participant, despite knowing the concept, was not able to draw it and evaluated the blend as equal or better than good (T5≥4). These results indicate that the system would be helpful to the user in 14 of the concepts (cases with a percentage of B>D higher or equal to the percentage of B<D), and its usefulness is particularly clear in eight (underlined in Table 12.1) from these concepts (36% of the 22).

When analysing the drawings made by the users, it is easy to observe how complex some of them are – e.g. drawings 1, 2 and 3 for *love song* in Fig. 12.3 were described by the users as "serenade", "writing a love song" and "sound waves". Yet, it is questionable whether these drawings would be perceived as *love song*. Moreover, some drawings could even be more closely related to other concepts – for *car factory*, 1 could be perceived as a driving car, and 4 and 5 could be interpreted as the icon for a garage for fixing cars. In example 2, the user even included the label "car factory" to make it perceivable. In most of these examples, the blend obtained better quality results than the drawing. Despite this, it is interesting to see how some of the drawings are very similar to the blends (e.g. the blend for *love song* in Fig. 12.1 and drawing 5 in Fig. 12.3). On the other hand, drawings 1 and 2 for *balancing act* were considered better than the blend, which shows that the system is not always capable of producing better solutions.

12.2.2 *Part #2 Perception*

After concluding part #1 of the experiment, we decided to focus on *perception* and *interpretation*. We conducted another study with the same participants and concepts of part #1.

12.2.2.1 *Experiment Setup*

We assigned a different group of concepts to each group of participants (i.e. user group 1 tested group 1 of concepts in part #1 and tested group 3 in part #2), which can be observed by comparing the number of tests (T) of each concept in Table 12.1 with the ones in Table 12.2. Each participant received a list of concepts and had to complete the following task for each concept:

- Identify the concept behind the blend.

For this task, an open-ended answer was expected and no indication was given in regards to the number of words to use, as we wanted

Table 12.2: Results of the naming study – number of tests conducted (T); number of tests with answer (A); for each word of the concept it is shown the count of tests in which the correct word was used (3), a similar term (2), a related term (1) and an incorrect term (0); the percentage of tests for five categories of (*word*1,*word*2) values, in which (3,3) is an exact match and (0,0) is an incorrect answer. Best results (highest percentages in two leftmost categories) are marked with ⋆.

			*word*1				*word*2					(2,≥2)	(1,≥1)	(0,≥0)	
	T	A	3	2	1	0	3	2	1	0	(3,3)	(≥2,2)	(≥1,1)	(≥0,0)	(0,0)
flag carrier	19	19	0	0	0	19	0	0	0	19	0,0%	0,0%	0,0%	0,0%	100,0%
growth rate	19	17	2	2	1	12	0	0	0	17	0,0%	0,0%	0,0%	29,4%	70,6%
packaging product	19	18	0	0	15	3	0	0	0	18	0,0%	0,0%	0,0%	83,3%	16,7%
⋆ *peace accord*	19	19	13	0	1	5	2	4	4	9	10,5%	21,1%	5,3%	52,6%	10,5%
power difficulty	19	19	1	3	7	8	0	0	2	17	0,0%	0,0%	10,5%	47,4%	42,1%
risk disclosure	19	18	0	0	2	16	0	0	15	3	0,0%	0,0%	11,1%	72,2%	16,7%
⋆ *security house*	19	19	9	2	8	0	6	3	2	8	10,5%	21,1%	26,3%	42,1%	0,0%
balancing act	22	18	0	0	0	18	0	0	12	6	0,0%	0,0%	0,0%	66,7%	33,3%
cigarette market	22	22	1	10	6	5	1	0	10	11	4,5%	0,0%	40,9%	36,4%	18,2%
failure risk	22	19	0	0	0	19	2	0	10	7	0,0%	0,0%	0,0%	63,2%	36,8%
future power	22	8	1	0	0	7	1	1	6	0	12,5%	0,0%	0,0%	87,5%	0,0%
⋆ *love song*	22	21	6	0	12	3	16	1	1	3	28,6%	0,0%	42,9%	28,6%	0,0%
⋆ *plane crash*	22	19	6	6	4	3	6	7	4	2	26,3%	26,3%	21,1%	26,3%	0,0%
vehicle operation	22	22	0	7	5	10	0	0	15	7	0,0%	0,0%	22,7%	77,3%	0,0%
business information	15	10	0	3	2	5	0	1	1	8	0,0%	10,0%	0,0%	50,0%	40,0%
⋆ *car factory*	15	15	5	0	1	9	3	1	10	1	20,0%	6,7%	13,3%	53,3%	6,7%
health risk	15	15	2	0	8	5	3	2	9	1	0,0%	6,7%	60,0%	26,7%	6,7%
market depression	15	11	0	0	0	11	0	0	5	6	0,0%	0,0%	0,0%	45,5%	54,5%
risk assessment	15	14	0	0	0	14	0	0	0	14	0,0%	0,0%	0,0%	0,0%	100,0%
rumor control	15	15	0	0	0	15	0	0	5	10	0,0%	0,0%	0,0%	33,3%	66,7%
sugar harvest	15	12	1	6	0	5	0	0	7	5	0,0%	0,0%	58,3%	0,0%	41,7%
university center	15	14	5	1	5	3	0	0	0	14	0,0%	0,0%	0,0%	78,6%	21,4%

Note: the values from *love song* are different from (Cunha et al., 2020b) due to correction.

to assess how the participants interpreted the blend without any constraint. For the same reason, participants were allowed to answer in Portuguese and English.

12.2.2.2 *Results*

A total of 407 tests were conducted, from which no answer was given in 43. A minimum of one word was used in the naming, a maximum of eight, mode of 1, median of 2 and standard deviation of 1.093. We conducted two different analyses to the results obtained.

The first one concerned the comparison of the name assigned by the participant and the concept used to produce the blend. The results of this analysis can be seen in Table 12.2. First, we analysed the name (see examples of answers in Table 12.3) and assigned a value to each of the

Table 12.3: Examples of answers given by participants in the naming study. We show answers that are similar to the concept and answers that are different. For each answer, we present the value assigned to each word of the concept, based on the quality of the answer – e.g. (3,3) is an exact match with the concept and (0,0) is a fully incorrect answer.

	similar		**different**	
	answer	value	answer	value
flag carrier	-		passport control	(0,0)
growth rate	time growth	(3,0)	hospital time	(0,0)
packaging product	gift wrap	(1,0)	privileged	(0,0)
peace accord	peaceful agreement	(3,2)	friendship	(0,1)
power difficulty	energy loss	(2,1)	sick planet	(0,0)
risk disclosure	searching error	(1,1)	blind	(0,0)
security house	home safety	(2,2)	locked	(1,0)
balancing act	-		cool theatre	(0,1)
cigarette market	smoking population	(1,1)	smoking risks	(1,0)
failure risk	death risk*	(0,3)	toxic	(0,0)
future power	metaphysical energy	(0,2)	electricity*	(0,1)
love song	favourite songs	(1,3)	heart beat	(1,0)
plane crash	plane accident*	(3,2)	crash	(0,3)
vehicle operation	car maintenance	(2,1)	hospital	(0,1)
business information	press work	(0,2)	fax	(0,0)
car factory	industrial car	(3,1)	fuel company	(1,1)
health risk	pharmacy risk	(1,3)	death (morte)	(0,1)
market depression	chinese crying	(0,1)	baby boom	(0,0)
risk assessment	-		flying money	(0,0)
rumor control	-		security	(0,1)
sugar harvest	cotton candy	(2,1)	reward growth	(0,0)
university center	university	(3,0)	med school	(1,0)

words of the blend concept: 3 if it was an exact match, 2 if the word used is a synonym ("home" for "house"), 1 if the word used is related ("musical" for "song") and 0 if it is completely different or inexistent. One example, for the blend *growth rate* the "time growth" is assigned with the scores (3,0) as the word "growth" is an exact match but "time" has little relation to "rate". It is important to mention that for answers in Portuguese these scores were assigned based on our translation. This process is not without issues, especially when it comes to language differences, e.g. in Portuguese the same word can be used for "home" and "house", resulting in a score of 3, but in English it would result in a score of 2. Some participants only used one word in the naming, which was often a clear reference to one of the emoji used in the blend. This task involved assessing if the participants could identify the two inputs of the blend based on the name assigned by them, and is in line with the *Unpacking Optimality Principle* – related to the easiness of reconstructing the inputs and the network of connections from the blend (see Section 4.1.2).

In general, most of the blends were not well perceived by the participants – 13 concepts had more than 89% of the tests with no answer or a score of 0 in at least one of the words, from which nine reached 100%. This number increases to 17 concepts (13 reaching 100%) if we also consider tests with a score of 1 in at least one of the words. The remaining five concepts were well perceived (scores ≥2) by more than 25% of the participants – 52.63% for "plane crash", 31.58% for "peace accord" and "security house", 28.57% for "love song", and 26.67% for "car factory". It is worth mentioning that for one of the concepts ("future power") only eight in 22 participants were able to provide a name. For some concepts, the bad results were due to one of the words – e.g "university center" none of the participants was able to fully perceive the concept but by observing Table 12.2 we can see that this was in great part due to the word "center", which no participant identified.

From analysing the blends, it is possible to conclude that there are different kinds of mapping between emoji and concept. For example, "car factory" is the only blend with a literal mapping – i.e. "car" is represented by a car emoji and "factory" by a factory one. Other concepts ("security house", "cigarette market", "plane crash" and "vehicle operation") are partially literal (only one word has a literal mapping) – e.g. the blend for "cigarette market" uses a cigarette emoji (Fig 12.1). Some blends ("love song" and "peace accord") use well-known symbols for both words – e.g. "love" being represented by a heart emoji (Fig 12.1) – and others ("university center" and "sugar harvest") use only in one of the words – e.g. using a candy emoji to represent "sugar"(Fig 12.1). From these nine concepts with less figurative word-emoji mapping, only three ("vehicle operation", "university center" and "sugar harvest") are not included in the set of nine concepts with high understanding rates (Table 12.1). Interestingly, the five blends with best results in naming (marked with ⋆ in Table 12.2) are among these concepts, which indicates that blends are easier to interpret when there is a more close connection between the words and the emoji. All the other concepts use more non-literal mappings – e.g. a crying face to represent "depression" in "market depression" (see Fig 12.1).

Following these findings, we conducted a second analysis of the results of this experiment, focusing on the connection between the emoji used in the blend (Fig 12.1) and the name assigned by the user. Our goal was to study how the user perceives the blend and its constituent emoji. In our analysis, emoji 1 is the emoji used to represent the word 1 and emoji 2 is the one used for word 2 – e.g. the house emoji in "security house" is used to represent word 2. For each test, we sought to identify the type of mapping done by the user for each emoji (literal vs non-literal, figurative or symbolic or symbolic). In some cases, an emoji is not represented in the name (e.g. giving the name "crash" for the "plane crash" blend indicates that there was an interpretation of the explosion emoji but no use of the plane emoji) or an unrelated

Table 12.4: Results of the perception study – for each emoji used in the blend (i.e. word 1 matches emoji 1) it is shown the percentage of tests with answer in which the participant had a literal interpretation (L) – e.g. hourglass emoji interpreted as "hourglass"– a non-literal interpretation (N) – e.g. hourglass emoji interpreted as "time" – and an unclassifiable interpretation (-), i.e. missing or unrelated.

	emoji 1			**emoji 2**		
	L	N	-	L	N	-
flag carrier	0,0%	47,4%	52,6%	57,9%	10,5%	31,6%
growth rate	5,9%	58,8%	35,3%	5,9%	52,9%	41,2%
packaging product	5,6%	55,6%	38,9%	83,3%	0,0%	16,7%
peace accord	0,0%	68,4%	31,6%	5,3%	52,6%	42,1%
power difficulty	5,3%	52,6%	42,1%	0,0%	47,4%	52,6%
risk disclosure	0,0%	38,9%	61,1%	5,6%	83,3%	11,1%
security house	0,0%	100,0%	0,0%	42,1%	15,8%	42,1%
balancing act	0,0%	22,2%	77,8%	0,0%	94,4%	5,6%
cigarette market	4,5%	77,3%	18,2%	0,0%	68,2%	31,8%
failure risk	0,0%	21,1%	78,9%	0,0%	89,5%	10,5%
future power	12,5%	25,0%	62,5%	12,5%	87,5%	0,0%
love song	19,0%	66,7%	14,3%	0,0%	100,0%	0,0%
plane crash	52,6%	31,6%	15,8%	15,8%	68,4%	15,8%
vehicle operation	31,8%	22,7%	45,5%	50,0%	31,8%	18,2%
business information	0,0%	50,0%	50,0%	20,0%	20,0%	60,0%
car factory	33,3%	6,7%	60,0%	20,0%	73,3%	6,7%
health risk	0,0%	66,7%	33,3%	0,0%	86,7%	13,3%
market depression	0,0%	0,0%	100,0%	36,4%	9,1%	54,5%
risk assessment	0,0%	14,3%	85,7%	50,0%	35,7%	14,3%
rumor control	0,0%	13,3%	86,7%	46,7%	33,3%	20,0%
sugar harvest	16,7%	50,0%	33,3%	41,7%	25,0%	33,3%
university center	0,0%	71,4%	28,6%	0,0%	0,0%	100,0%

interpretation occurs (e.g. the name "blind" for the blend "risk disclosure"). The results can be observed in Table 12.4.

In the majority of the cases, the interpretation is non-literal for the majority of the participants. However, for some emoji, there is a clear tendency for a literal interpretation: emoji 2 of "flag carrier" ("passport control" emoji) is often interpreted as a police officer; emoji 1 of "packaging product" ("wrapped gift" emoji) is interpreted as "gift" by 15 out of the 18 participants who provided a name; and emoji 1 of "plane crash" ("airplane arrival" emoji) is interpreted literally by 52.6% of the participants. Such tendency is advantageous when the link emoji-word is also literal (e.g. plane in "plane crash") but it is troublesome when this connection is more non-literal (e.g. in "packaging product" the gift is used to represent "product").

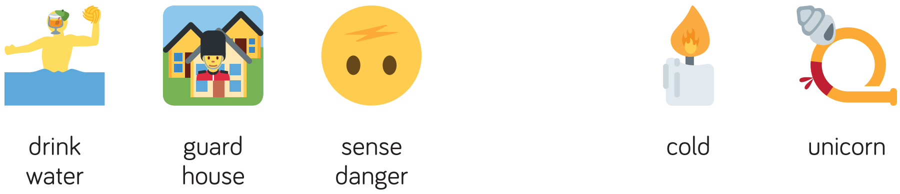


Figure 12.4: Three blends for *dog* using different related concepts (on the left), and blends for *cold* and *unicorn* (on the right).

For 24 out of the 44 emoji used in the blends we were not able to classify the interpretation of at least a third of the participants (either a missing or an unrelated reference, values in column "-" in Table 12.4).

12.3 OVERALL ANALYSIS OF EMOJINATING

In this part of the thesis, we described the implementation and evaluation of *Emojinating* – a system that presents the user with visual representations of concepts using emoji and emoji blending. First of all, it is important to call the reader's attention to the fact that our intention is not to generate emoji – we are well aware of emoji requirements in terms of clarity and we do not want to further increase interpretation issues such as the ones identified by Miller et al. (2016). Our main goal is to produce visual representations of concepts, using emoji as a means and not an end-goal.

Overall, we consider that the representations obtained with *Emojinating* are visually and conceptually interesting and often unexpected (in a positive way), which is supported by the results of the presented studies (e.g. the results obtained in the surprise assessment task of the study described in Section 9.4.2).

The system is able to generate a wide variety of results, both with the same emoji – e.g. *rain man* in Fig. 8.10 – and with different ones – e.g. *dog* in Fig. 12.4 (left). The visual blending process produces blends that represent the input concept and vary in terms of the degree of conceptual complexity – the blends for *dog* (Fig. 12.4) are harder to understand than the one for *man bat* (Fig. 12.6). Moreover, the system is able to make indirect connections, e.g. *wine polo* has a literal representation whereas the one for *car* is non-literal (Fig. 12.5), which can be interpreted as "car is a fast way to die".

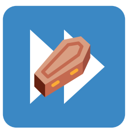


Figure 12.5: Blend produced for "car"

On the other hand, results are not always easy to understand. An example of this is the set of results obtained when introducing the concept *dog* (see Fig. 12.4). To propose blends for the initial concept, the system makes connections to other concepts. In the case of *dog*, the related concepts are: *drink water*, *guard house* and *sense danger*. Even though all these make sense for describing *dog*, it is not easy to perceive *dog* just by looking at them.

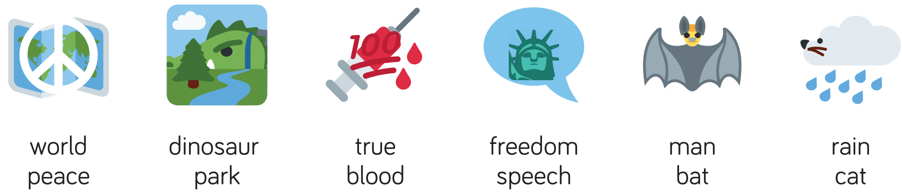


Figure 12.6: Examples of visual blends of double-word concepts

However, blends do not always make sense and, therefore, cannot in some cases be considered representations for the introduced concept – e.g. in one of the initial experiments of *Emojinating*, the concept *unicorn* was extended to *spiral horn*, which then leads to a shell emoji (for *spiral*) and a postal horn emoji (for *horn*), and results in the blend shown in Fig. 12.4 (right). In other cases, the search for related concepts even leads to opposite meanings. This results in the generation of blends that do not represent the introduced concept but something that represents its opposite instead. One example of this is the blend for the concept *cold*, in which a candle is represented (Fig. 12.4).

There is no doubt that the performance of the system is dependent on the input emoji set and the semantic knowledge associated with it. As such, it might generate interesting blends for some concepts and uninteresting for others. Moreover, by using the *ConceptNet* Application Programming Interface (API) we are also subject to updates on their end, which may cause changes in the results of the system.

The main question that arises is whether visual blending can be considered suitable for concept representation in the case of *Emojinating* (in line with research question A). The high coverage of the core concept list is an argument in favour (see Section 8.2). However, when analysing the usage of the system by participants some issues point otherwise. First, the occurrence of *emoji hiding* – one of the emoji was partially or even totally hidden – which is an exploit of the system and does not make usage of visual blending (see Section 10.4.2). Another unfavourable result is the high occurrence of *juxtaposition* (Table 11.6), which we consider as a poor type of blending, having little gain when compared to a sequential positioning approach. Despite this, there are certain cases in which *juxtaposition* has clear advantages (e.g. the "dumpster fire" example given in Section 7.6.2).

On the other hand, the usage of *replacement* is a good indication that the visual blending is useful. When analysing user drawings (obtained in a *Pictionary-like* task), users tend to draw existing objects and use a juxtaposition-based approach. Replacement often leads to non-literal solutions, which normally require more complex reasoning from humans. As such, the system provides a quick way to present the user with solutions that require such reasoning. These topics are indication that different blend types have different advantages.

The two advantages of the system are that it provides the user with the possibility of choosing between different blend type solutions, often leading to more than one solution deemed good and that it follows a multi-purpose approach, allowing the user to introduce any concept without requiring changes to the configuration or extra data.

As far as *co-creativity* is concerned, one of the most used arguments in favour of co-creative systems is their potential to foster users' creativity (Karimi et al., 2018b; Liapis et al., 2016). The interaction with the system allows the user to evolve solutions that match their preferences and, at the same time, both the user and the system are constantly influencing the perception of one another, leading to novel ideas. The results obtained in the study described in Section 12.2 provide evidence that this interaction leads to better solutions than the ones drawn by the user alone. The potential of the system is even clearer if we consider that in two cases participants who knew the concept were not able to draw it and afterwards considered the blend as a good representation.

One last aspect that points towards the high complexity of using visual blending for concept representation concerns issues of perception. Even though good results were obtained in the comparison of blends with drawings, when assessing whether participants could perceive the concept behind blends, results pointed in the opposite direction – only five out of 22 concepts were well perceived by more than 25% of the participants. This means that, despite being considered good representations by their creators and participants who knew the concepts behind the blends, people with no prior information could not guess the concepts in most of the cases. At first, one would say that this indicates that visual blending is not adequate for concept representation. However, when analysing drawings by users, it is easy to point out that most of them are visually more complex than most blends, some of the drawings are non-literal and make use of a related concepts, e.g. drawing a serenade to represent "love song". Furthermore, upon asking participants to draw representations we also asked them to provide a description of their drawing. On a quick analysis, we notice that some of the descriptions are long, which points towards complexity of representation. Based on these facts, we believe that issues of interpretation do not entirely rely on the efficiency of visual blending but also on the easiness of representation – e.g. representing "risk disclosure" is much harder than "love song". As such, we anticipate that the interpretation of drawings would also be a difficult task. In order to further analyse the perception and interpretation of drawings and blends, more studies are necessary.

12.4 SUMMARY

In this chapter, we summarised the development of *Emojinating*, highlighting some of the differences between versions. Then, we focused on two topics related to the production of visual blends: *usefulness* and *perception*. We conducted a user study to address these topics and the results indicate that blends are easier to interpret when there is a closer connection between the words and the emoji. Moreover, the results showed that the perception of blends is not always easy but that there are benefits of using *Emojinating* to produce visual representations of concepts, helping in ideation and creativity fostering. We ended the chapter with an overall analysis of *Emojinating*, discussing some of its key aspects, especially in regards to the representation of concepts.

Part IV

GLYPHS, FLAGS AND OTHER EXPLORATIONS

The visual representation of concepts is an open-ended and multi-purpose task. So far, we have mostly devoted our attention to the generation symbols based on the combination of two words. In this part of the thesis, we explore other applications that are based on the work that we presented in the previous chapters. We start by describing an approach to use the emoji retrieval component of *Emojinating* in the context of information visualisation. Then, we change our focus from icon-like symbols to another type of symbol: flags. We describe a system that produces flags that represent trending topics inferred from news sources.

13 STRATEGY FOR DATA GLYPH GENERATION

In the previous part of the thesis, we have described *Emojinating*, a system that uses visual blending of emoji to visually represent concepts. In this chapter, we address the potential application of part of *Emojinating*'s approach for visualisation purposes.

In this chapter, we outline an approach based on the emoji searcher component (described in Section 8.1.2) to automatically generate data-related glyphs. First, we highlight how emoji are suitable for data visualisation due to their conceptual coverage and to the existence of Scalable Vector Graphics (SVG) emoji datasets. Then, we assess the performance of our approach by comparing generated glyphs with existing ones and estimate its usefulness in representing datasets used by other authors.

The approach described in this chapter is only partially implemented, as it is based on *Emojinating*'s components. Nonetheless, our main goal is to demonstrate the potential of the approach, rather than to have a fully functional implementation.

This chapter is based on the work presented in the paper by Cunha et al. (2018).

13.1 CONTEXT

In the context of information visualisation, data glyphs are used for the representation of multidimensional data (Chernoff, 1973). Glyphs can be described as composite graphical objects that use their visual and geometric attributes to encode multidimensional data by mapping each dimension of data point to the marks of a glyph (Anderson, 1957; Wittenbrink, Pang, and Lodha, 1996).

There are different kinds of data glyphs with varying designs and conceptual diversity. Some glyphs are abstract, e.g. *polygon glyph* (Li, Li, and Zhang, 2015), and others have an iconic nature, e.g. *faces*, *cars*, or even *flowers*. When considering iconic glyphs, they can be unrelated to the data thematic (e.g. a *face glyph* representing *forest fires* data) or be related in a literal (e.g. a *face glyph* representing data on *facial features*) or a figurative way (e.g. a *face glyph* representing *non-facial anthropometric* data). Also, a variety of surveys about data glyphs and their usage have been published in the recent years (Borgo et al., 2013; Fuchs et al., 2017).

In this chapter, we base our work on one type of glyph: *Chernoff faces* (Chernoff, 1973). This type of glyphs encodes multidimensional data using facial features such as the length of the nose, the orientation of the eyebrows, the shape of a mouth, among others. One particular feature of *Chernoff faces*, which originated other alike glyph designs, is its resemblance with a human face. Although this kind of glyphs performs poorly in terms of response time, as well as the accuracy of glyph decoding, when compared to other existing glyphs (Lee, Reilly, and Butavicius, 2003), the metaphoric projection and analogy with faces make *Chernoff faces* a powerful tool for conveying complex data.

In general, using glyphs related to the data is said to have perceptual advantages, which leads to easier interpretation and better accuracy (Siirtola, 2005). In addition, some authors justify the usage of certain glyph designs with reasons related to human aptitudes, such as the ability to easily recognise faces, e.g. *face glyphs* (Chernoff, 1973), or to visually discriminate natural shapes, e.g. *leaf glyphs* (Fuchs, Weiler, and Schreck, 2015). For these reasons, several authors not only argue in favour of data-related glyphs but also point out that it would be advantageous for a glyph-based visualisation tool to have different types of glyphs, which could be chosen by the user and allow a better match to the data (e.g. Siirtola, 2005).

Such a tool is normally considered difficult to implement, as it would require a large repository of glyphs prepared for data representation. Considering repositories such as image banks falls short, as images, due to their pixel-based nature, are often unsuitable for variation according to data. These requirements partially align with the ones identified for concept visual representation (described in Section 4.3).

On the other hand, we believe that emoji have several properties which make them adequate for this task. We previously analysed these properties in Section 7.6.3. As such, we describe an approach to use emoji in a visualisation tool.

13.2 RELATED WORK

In the 1990s and the beginning of the 2000s there was increasing interest in intelligent or "smart" graphics, usually referred to as *automatic visualisation* (Casner, 1991; Iizuka et al., 1998; Petajan et al., 1997). This approach mostly consists of a rule-based mapping between graphical elements and data type accompanied with additional underlying mathematical statistics, which are used to summarise the data. Automatic visualisation is efficient to get a first impression of the data. However, the disadvantage of this approach is the fact that automatic visualisation is just statistical projections on visualisation artefacts. It is extremely difficult to extract any insight or high-level information from such projections.

One example of automatic visualisation is *AutoVis* tool, developed by Graham and Leland (Wills and Wilkinson, 2010). Given the dataset, the system decides how to appropriately visualise it based solely on visualisation theory and without presupposing a predefined visualisation model. The overall strategy for selecting proper graphical elements is based on the *Grammar of Graphics* introduced by Wilkinson (2006). The system uses a prioritising mechanism to select the most "interesting" views to display, which is similar to *Google*'s Page Rank. Finally, the visualisation utilises statistical methods to find relationships in the given data, and represents it using a graph model, accompanied with simple graphics (e.g. area charts, bar charts, box-plots, etc.) that summarise computed statistics.

Another example of automatic visualisation is the work of Mackinlay, Hanrahan, and Stolte (2007). The tool, called *Show Me*, integrates a set of user interface commands that enable the automatic generation of tables of views for multiple fields in the dataset. Provided with the specified rows and columns the tool automatically selects *Mark Type* – graphical element (e.g. bar, line, shape, etc.) – and *View Type* – graphical methods to represent data (e.g. scatter plot, bar chart, etc.). The decision is made based on the predefined rules, which result from – "best ways of producing charts and graphs". Finally, each additional command can yield a particular type of view. Likewise, our approach is intended for the production of data glyphs based on the query that a user performs, or based on the data, in particular meta-data, provided to the system.

13.3 approach

This work builds upon the previously described *Emojinating* system (Section 8), in which emoji are automatically retrieved to be used to visually represent concepts introduced by a user. As described in Section 8, *Emojinating* has three components: (i) *Concept Extender* (uses *ConceptNet* to search for related concepts to a given one), (ii) *Emoji Searcher* (searches for existing emoji semantically related to a given single-word or double-word concept), and (iii) *Emoji Blender* (produces blends by visually blending emoji). In the context of this chapter, we use the *Emoji Searcher* and the *Concept Extender*.

The *Emoji Searcher* component takes words as input and retrieves emoji that are semantically related to them, using semantic data from *EmojiNet* (Wijeratne et al., 2017b). By using these two components together, we are able to obtain emoji that are directly related to an introduced word – emoji that matches the word – or indirectly related – emoji that match words that are related to the introduced one, retrieved by the *Concept Extender* component. This allows the retrieval of emoji that can be used as glyphs for an introduced thematic.

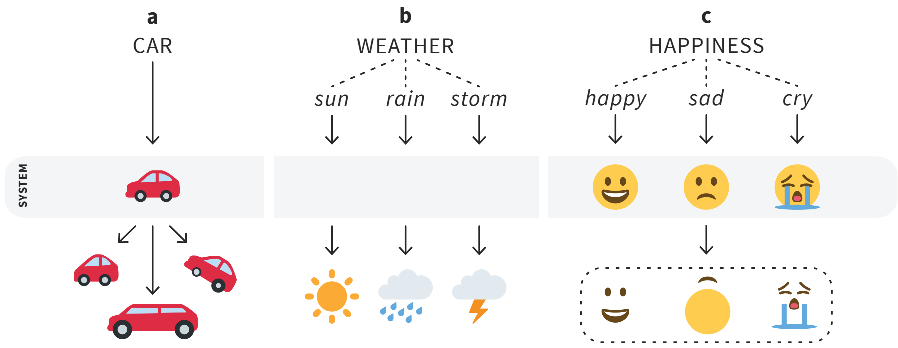


Figure 13.1: Different functioning modes

13.3.1 *Functioning Modes*

Emoji can be used as data glyphs in different ways, depending on data type, more specifically on the semantics of given data. We considered three functioning modes (see Fig. 13.1):

a) Glyph as single emoji: the emoji is used as a glyph and the individual shapes, in which it can be decomposed, are seen as visual variables. In such cases, the user obtains possibilities of glyphs for the introduced thematic (e.g. *car* thematic on the left side of Fig. 13.1). Similar to *Chernoff faces*, in this mode the applied data variables can be both categorical or numerical;

b) Glyph as a set of emoji: the user does not introduce a single thematic but several categories. Emoji are retrieved for each category and they are used as a set. One example is a weather map, in which each type of weather is represented by a different icon (e.g. sunny, rainy, etc.). This mode is intended to be related to categorical data;

c) Glyph as a combination of emoji parts: similar to *mode b* but with the difference that the emoji for each category are merged into a single glyph, and the visual variables are not used in a continuous way, but in a combinatory one. In such cases, there is a high amount of shared features among the emoji for each category. The parts that change are identified and separated (e.g. in Fig. 13.1 for the *happy* category only the mouth is different from all the other emoji).

In all three modes, the user introduces a topic and receives emoji that represent it. These emoji are then transformed (*mode a*), used by replacement (*mode b*) or combined (*mode c*). The modes are related to the type of data: *mode a* will be used for numerical data, whereas for situations in which there is only categorical data *modes b* or *c* will most likely be preferred. In *mode a* categorical visual variables may consist, for example, in colour variation (e.g. the colour of a car glyph).

13.3.2 *Architecture*

Even though our purpose is to only demonstrate the potential application of our approach in Information Visualisation, we describe the architecture of the system. The proposed approach would follow a four-step pipeline:

1. **Topic identification:** this step consists in the identification of the topic to be searched. The topic is provided by the user and is used to gather emoji to be glyph candidates. Depending on the type of data, the introduced topic may be a thematic (e.g. functioning *mode a*, described in subsection 13.3.1) or categories of the data itself (e.g. functioning modes *b* and *c*).

2. **Glyph generation:** after gathering candidates to be glyphs, these are filtered (removing inadequate ones) and the remaining ones are prepared to be used in the visualisation. This step can be divided into the following tasks: (i) emoji deconstruction, (ii) identification of visual variables.

3. **Configuration by the user:** the system presents the user with glyphs, their possible visual variables and suggested variation limits. Then the user is able to configure the assignment of visual variables, as well as, establish the variation limits, with the help of the suggestions of the system.

4. **Setup of the display:** configuring the final view of the visualisation, i.e. how the glyphs should be organised. Two examples of this are a grid layout or positioning according to a given parameter (e.g. positioning glyphs in a map according to geographic locations).

13.4 DISCUSSION

This section provides a discussion of the possible outcomes of the approach. The analysis is divided into different sections: (i) comparison with existing data glyph types (*faces*, *cars* and *flowers*), (ii) results for dataset thematics used by other authors, and (iii) open issues.

In order to analyse the potential of our approach in terms of usefulness in Information Visualisation, we compared it with existing glyph techniques. To do this, we conducted a bibliographic research that used the systematic review on glyphs by Fuchs et al. (2017) as a starting point (later extended to other papers). We focused our search on iconic data glyphs and collected a total of 40 research papers. These were analysed and 13 were discarded as we considered that the glyph used was too abstract. For the final selection, we collected the following information: type of glyph(s) used (e.g. car), number of total glyph visual variables, number of visual variables used, thematic of the dataset

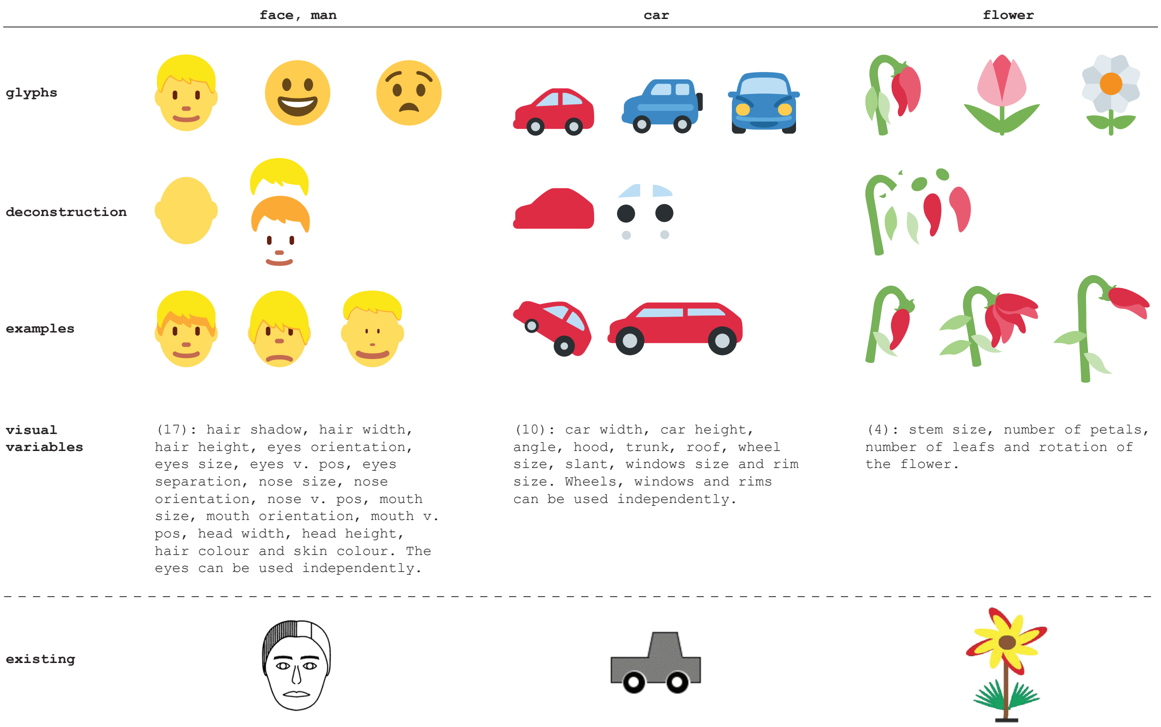


Figure 13.2: Examples of data glyphs obtained with our system, their deconstruction, usage examples, list of possible visual variables and comparison with glyphs used by other authors – face (Flury and Riedwyl, 1981), car (Siirtola, 2005) and flower (Chau, 2011).

used and data variables represented. In this chapter, we present some examples to illustrate our study.

In current approaches, data glyphs are either used of-the-shelf or custom made for the visualisation at hand. In comparison, our approach aims at providing a multipurpose way of generating glyphs, without previously defined thematics.

On a general level, this approach allows the user to get emoji that represent the introduced topic. It would lead to a system that can provide the user with several emoji options that are easy to prepare for a visualisation task, due to their layered structure. The system can be implemented as semi-automatic, by making the user responsible for configuring the visualisation.

The approach described in this chapter is partly implemented as it is based on components from *Emojinating*, with which it is possible to retrieve emoji for introduced words. The examples of glyphs presented in this chapter were automatically gathered using the system.

13.4.1 *Comparison with existing glyphs*

The first analysis that we consider important to be performed has to do with the generation of glyph types that have been often used in Information Visualisation. In our bibliographic research focused on iconic glyphs, we were able to identify three glyph types that we considered as benchmarks: *face*, *car* and *flower*. We used our system to obtain emoji that were similar to these glyph types and compared them (see Fig. 13.2).

The analysis of the glyphs produced by our system was focused on overall visual representation, the number of visual variables and the easiness of variation. The number of possible visual variables can be estimated by decomposing the emoji into its individual shapes. Despite that, in some cases, the shape can be partially hidden or too small to apply transformations in a perceptible way and without influencing other shapes. Each shape can vary in terms of *location*, *size*, *orientation*, *colour* (*hue*, *value* and *saturation*), *texture*, among others. Obviously, some shapes may be more suitable for some of the transformations – e.g. a *colour* or *texture* variation in a very small shape will most likely have a very reduced effect on the perception of change. We estimated the number of visual variables for each of the glyphs obtained. The comparison with existing glyphs addressed two topics: visual appearance of the glyph and the number of visual variables.

13.4.1.1 *Faces*

There are different versions of the face glyph, which vary in terms of number of visual variables – from four in the work from Nelson et al. (1997) to 32 in the work from Flury and Riedwyl (1981) – and in terms of realism. Our system is able to produce different candidates to be used as face glyph, which visual differ among each other (e.g. not all have hair or the inner shape of the eye). We are able to produce a face glyph with 17 visual variables. Despite this value being very distant from the one from *Flury-Riedwyl face* glyph, which has 32 (Flury and Riedwyl, 1981), in the majority of the analysed papers that employ face glyphs less than 15 visual variables were necessary to represent the data.

13.4.1.2 *Cars*

Concerning the car glyph, we are able to produce emoji with as many visual variables (10) as the ones from existing car glyphs (Siirtola, 2005). We are also able to present the user with different versions (i.e. different types of cars), which have a higher number of visual variables. When comparing the appearance of our versions with the one from Siirtola (2005), ours can be considered more realistic.

13.4.1.3 *Flowers*

Two different versions of the flower glyph currently exist: one with a stem (3-4 visual variables) (Chau, 2011; Zhu, 2002) and one with just a flower shape (3 visual variables) (Li, Li, and Zhang, 2015). Both implementations use functions that automatically produce any number of petals. As our approach uses emoji previously drawn, such visual variation would require a custom implementation. Despite that, we are able to match the number of visual variables of existing flower glyphs.

Overall, we are able to obtain a similar number of visual variables to the ones from existing glyphs and we consider our versions more visually appealing. Moreover, whereas implementation using functions allows greater flexibility, our approach allows a higher variability of glyphs.

13.4.2 *Thematic representation*

In our investigation, we collected the thematics of all the datasets represented by the analysed papers (Fig. 13.3) and used our system to produce glyph candidates for them. This allowed us to assess the performance of the system in suggesting data-related glyph candidates.

The majority of the analysed papers used unrelated or metaphoric glyphs to represent the data, e.g. representing fires with a leaf (Fuchs, Weiler, and Schreck, 2015). With our system, we were able to obtain data-related candidates for all the thematics, ranging from literal (e.g. a fire icon to represent fires) to *metaphoric* (e.g. a magnifying glass to represent *Google* search results).

The system is able to generate several possible glyphs for the same thematic. For example, for the thematic *fire*, we obtained four different candidates – going from *literal* (flame) to *non-literal* (fire truck) – with a different number of possible visual variables.

13.4.3 *Overall Analysis and Open Issues*

The glyph candidates obtained with the system vary visually and in terms of structure (e.g. face has specific locations for the eyes, mouth and nose; a flower has a different behaviour related to the positioning of petals). This leads to some issues that require further analysis.

13.4.3.1 *Transformation*

This structural difference affects the way transformations are applied. As already mentioned, whereas current glyph techniques use functions in glyph construction, we use previously designed emoji and the transformations are mainly done using distortion of existing shapes. As such, basic geometric transformations like scaling and translation are

glyph	word	thematic
	fire	Forest Fire [Fuchs 2015]
	education	Investment in education [Li et al. 2015]
	kill	Crime [Nelson et al. 1997]
	music	Classical Music [Chan et al. 2010]
	note	Swiss-banknote [Hamner et al. 1987]
	nut/seed	Seeds [Fuchs 2015]
	patient	Psychiatric patients [Mezzich & Worthington 1978]
	runner	Marathon runners [MacGregor & Slovic 1986]
	search	Google search results [Chau 2011]
	student	Investment in education [Li et al. 2015]
	temperature	Temperature / pressure [Casey 1987]
	toxic	Toxic substances [Brown 1985]
	twin	Twins [Flury & Riedwyl 1981]

Figure 13.3: Glyphs obtained for dataset thematics used by other authors (Brown, 1985; Casey, 1987; Chan, Qu, and Mak, 2010; Chau, 2011; Flury and Riedwyl, 1981; Fuchs, Weiler, and Schreck, 2015; Hamner, Turner, and Young, 1987; Li, Li, and Zhang, 2015; MacGregor and Slovic, 1986; Nelson et al., 1997).

easy to apply but complex transformations like duplication of shapes (e.g. flower petals) require custom implementation.

Some shapes may be more suitable for some of the transformations than others. These aspects should ideally be analysed for each glyph candidate and should be taken into account in order to define transformation rules to automatically suggest variation types and ranges.

13.4.3.2 *Variation limits*

Ideally, the system should be able to automatically assess and suggest suitable ranges for each visual variable. To estimate these limits, it is necessary to develop methods for analysing shapes, identifying points of contact and overlaps.

13.4.3.3 *Salience and Complexity*

Some glyph characteristics will be difficult to assess due to the high variety of results. One example is the salience difference among visual variables of a glyph, i.e. some elements are more important than others in perception (De Soete, 1986). Another issue has to do with the complexity degree: some glyphs are very simple and others are much more complex (e.g. in Fig. 13.3 *fire* is simple and *education* is complex). This makes it necessary to further study the impact of glyph complexity on perception and interpretation. The analysis of these characteristics would be impossible to do for every glyph and, as such, different approaches should be followed (e.g. assessing salience using pixel-based difference calculation and limiting the system to simpler emoji to avoid complexity).

Despite these open issues, in our opinion the major advantage of our approach is the automatic proposal of data-related glyphs, which we believe can be achieved with our approach. It is also important to mention that different emoji datasets exist, thus more glyph possibilities.

13.5 SUMMARY

In this chapter, we presented an approach that uses two of the components developed for *Emojinating* (the *Emoji Searcher* and the *Concept Extender*) in the context of Information Visualisation. The approach has the goal of presenting the user with emoji that can be exploited as data glyphs for data visualisation purposes. We demonstrate the potential of the approach by comparing existing glyphs with emoji that can be obtained for the same thematic.

14 FLAGS

This chapter introduces flags as symbols that are associated with multiple meanings. We start by describing related work and presenting the characteristics of flags that make them interesting for this thesis.

Then, we propose the possibility of using a flag to represent changes that occur in short timeframes. We present a system that generates flags based on trending topics of countries, retrieved from real-time news. These topics are used to drive a process of visual blending that alters the original flag of the country. In this sense, the produced flags can be seen as visual representations of the current "mood" of the country. We assess the impact of generated flags by conducting a user study, focused on perception and interpretation.

This chapter is based on the work presented in the papers by Cunha, Martins, and Machado (2020b) and Cunha et al. (2020c).

14.1 CONTEXT

Flags are among the symbols of a nation that help the formation and maintenance of a national identity (Elgenius, 2011), both internally – among its citizens – and externally – keeping a coherent sense of oneness in the perception by other countries and entities. This process of maintaining a collective identity is described as an "ongoing, dynamic process in which historical symbolic meanings are constantly recycled, actualised, challenged, renegotiated, and reconfirmed" (Geisler, 2005).

Changes that occur in a country throughout history are often reflected in the design of its flag, whose elements bear meaning and are part of the country's culture. In the past, the dissemination of these changes was slow and of limited access. As such, modifications to the design of a flag are normally sporadic and, in most cases, a flag remains unchanged for long periods.

By looking at country flags one can easily identify similarities among them, which point to how different flags influenced each other throughout history (Healy, 1994). The exploration of this relational character is observed in imaginary scenarios, for example, an alternate universe in which Nazi Germany and Imperial Japan won World War II, which is depicted in the Amazon's mini-series *The Man In The High Castle*, based on Philip K. Dick's novel of the same name (Dick, 1962). The series shows the design of fictional flags for an America ruled by Nazi and Japanese forces, which resemble the original flags (Heller, 2015).

Going beyond the reflection of its evolutionary path, a flag exerts its most significant role as a means of conveying the intended image of the entity that it stands for. One example is the design of a new European Union flag by Rem Koolhaas based on the essence of the European project as a joint effort of different nation-states, each with its own identity but together contributing to a plural identity of EU.[1] The redesign resulted in a barcode-style flag featuring the colours of EU countries, transmitting the idea of individual identities and simultaneous advantages of acting together.

On the other hand, the sense of identity also has fragilities. The value of one's identity makes it so that it is often prone to exploitation and manipulation, for example by the misappropriation of flags. The *Double Standards* project (Pater, 2012) investigated 59 seajacked ships that mask their owner's nationality by purchasing a "cheap flag" from another country to avoid taxes and environmental regulations. Such examples highlight how volatile an identity can be, especially in a time when individual identity loses power to the growing advances of globalisation. Moreover, in addition to this dissolution of individuality, in the current society characterised by constant change, the idea of an immutable identity becomes more and more questionable.

This sense of fluid identity is explored, for example, in the project *Net.flag* (Napier, 2002), which is based on the idea of an "ever-changing flag of the Internet" that anyone can alter upon visiting its website. In our opinion, these issues are ground for an important discussion on how the identity of a nation is represented by its flag and on the impact brought by changes in this national symbol.

Another interesting project involving changing flags is the brand identity for the *Westman Islands nation* (an invented hacker micro-nation) designed by Mariagloria Posani, Giulia Ponzetta and Emanuele Sciolto (Guida, 2014). For this identity, a "flexible" flag was designed to react to sound volume and tone variations, causing a series of pixels to rearrange, generating different versions.

Our society has now easy access to global information, which results in a sense of constant change. One can question whether a nation only possesses an identity or, based on this constant change, if it can also be assigned what we refer to as "mood" – i.e. what is happening in the country at the moment. This concept is aligned with a flag campaign[2] that was made for *Grande Reportagem Magazine* in 2005, in which the meanings of seven flags were changed based on shocking facts about the country. As such, the following question is posed:

Can the "mood" of a country be represented in its flag?

With this question, our goal is to explore how flags can be used to visually represent concepts.

1 oma.eu/projects/eu-barcode
2 creativecriminals.com/print/grande-reportagem/flags

14.2 RELATED WORK

Flags are normally custom-made and designed using elements that have meanings assigned to them. Nonetheless, more systematic strategies can also be used to produce flags. One strategy consists in generating flags from scratch using a pre-defined grammar.

For example, in the year 2000 Arjan Groot founded the organisation *Universal Authority for National Flag Registration* (UNFR) (Groot, 2000), which developed a flag coding system by analysing the colours and patterns of existing flags. In this system, a flag is composed of: (i) a background colour, (ii) a pattern in any of the other six colours, or a combination of patterns, and (iii) an optional symbol in one colour. This system not only indexed UN member countries but produced thousands of unclaimed flags. Another example is the web app *Scrontch's Flag Designer*[3] by Lars Ruoff, which allows the user to produce flags based on a grammar with three element categories: division (12 different types of flag division, for example three horizontal stripes); overlay, which controls the presence/absence of a shape element (the user can select from 11 different shapes, e.g. square on the top left quadrant); and symbol, which is positioned either centred on the flag or inside the overlay element (the user can select from a set of 17 different symbols, e.g. an eagle icon). In addition to the selection of the element for these three categories, the user can also select colours for them (8 different colours). It has also the option for random generation and export of the final result. Similarly, Whigham, Aldridge, and Lange (2009) defined a "flag language" – composed of basic elements (e.g. background) and functions (e.g. clipping) – and used an interactive evolutionary approach to produce new flags.

Another way of producing flags is by combining existing ones – i.e. *visual blending*. Examples of visual blending of flags are: the proposed EU flag by Rem Koolhaas,[4] which uses a barcode style featuring the colours of EU countries; the fictional flags designed within the context of *The Man In The High Castle* (Dick, 1962) by merging existing ones (Heller, 2015); or the combination of two flags using a masking technique to represent nationality deception by ships seajacked by Somali pirates[5] (Pater, 2012). There are several computational systems that use a visual blending approach to flag production. For example *Net.flag*, a project commissioned in 2002 by the *Guggenheim Museum*, is an online flag editor in which flags can be produced by removing or adding elements belonging to existing flags (Napier, 2002). It is presented as having an "ever-changing flag of the Internet", which any website visitor can alter. Similarly, the project *Atlas of Potential Nations: Computationally Designed Nations*, developed by *Emblemmatic*, produces

3 flag-designer.appspot.com/
4 oma.eu/projects/eu-barcode
5 www.doublestandardsofpiracy.org

names and flags for new nations by combining the existing flag elements. As the project website[6] describes, the system uses *Markov chains* and *context-free grammars*. In addition to these projects, there are also *Twitter* bots that generate flags – e.g. the *Flags Mashup Bot*[7] mixes existing flags by applying the colours of one flag to the elements of another; or the *FlagBot*[8] produces new flags by putting together elements of several existing flags and changing their colours.

From all these examples of flag production, none seems to explicitly explore what we consider the most relevant aspect in flag generation: the meaning of the flag. Nonetheless, it is clear how a flag can be used to encompass several meanings and how visual blending is prone to be used to produce new flags.

14.3 IMPLEMENTATION

Flags can be analysed in multiple ways – e.g. in terms of *complexity*, *colour*, *similarity*, among other criteria.[9] Regarding an analysis of a single flag, three aspects have a central role:

1. structure, i.e. how it is divided, what elements it includes, etc.;
2. meaning associated with its elements;
3. what the flag symbolises, e.g. a national flag represents a nation.

However, approaches to flag generation mostly focus on *structure* and give little attention to the other aspects. Our approach combines the three while giving special emphasis to the meaning of the flag elements, using it to change what the flag represents.

More specifically, our goal is to represent what we refer to as the "mood" of the country. The concept of "mood" is based on the expression *I'm in the mood for* [*something*], which is normally used in association with feelings that do not last long. To achieve this, we use the flag of a country as the starting point and apply changes according to real-time data about the country.

This approach leads to three main questions that need to be addressed: (i) "where to get the necessary data?", (ii) "how to apply changes to the flag in a way that they make sense?" and (iii) "how to present the results to the user?". In this section, we describe the implementation of a system called *Moody Flags* and we explain how we dealt with each of these questions.

6 emblemmatic.org/atlas/
7 twitter.com/FlagsMashupBot/
8 twitter.com/FlagBot1
9 flagstories.co

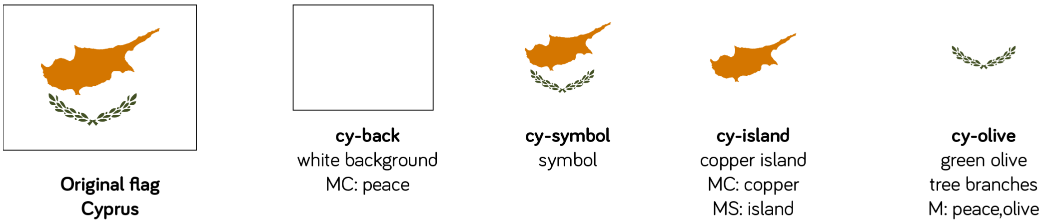


Figure 14.1: Example of the data collected for the Cyprus flag. The figure shows the id assigned (e.g. *cy-island*), a description ("copper island") and meanings (M stands for general meaning, MC for the meaning of colour and MS for the meaning of shape).

14.3.1 *Flag dataset*

The first issue to address had to do with obtaining the necessary data for flag generation. By searching existing projects on flags, we were able to find sources of three kinds of data: *visual*, e.g. a dataset of Scalable Vector Graphics (SVGs) of flags (*flag-icon-css*[10]); *semantic*, e.g. the *Net.flag*[11] project (Napier, 2002); and about *flag structure*, e.g. the platform *Flag Identifier*[12] (Sarajčić, 2007). Data on flag structure is very useful for generating new flags from scratch. In contrast, when producing flags by transforming existing ones, the most useful types of data are semantic and visual. Since we could not find any dataset that associated both types of data, we decided to produce one.

As starting point, we used version 3.3.0 of *flag-icon-css* SVG dataset, which contains 257 flags. However, image files of the original dataset were not properly structured nor had they proper layer identification. For this reason, we produced a new version of the dataset, in which we organised the layers into groups according to flag structure and assigned the ids to the layers. This allows not only the replacement of individual shapes but also groups. For each element of a flag, we collected meanings on colour, shape and overall meaning (see example in Fig. 14.1), from four main sources: the project *Net.flag*, the book *Complete Flags of the World* (Wills, 2008), *Wikipedia* flag pages and "Meaning of [...] flag" posts on *Reddit*.[13] This process mostly involved reducing long descriptive sentences into keywords. To establish a correspondence between visual and semantic data, we used the ids assigned to the layers of the SVG files. Due to its time-consuming character, the SVG structuring and meaning collection is still an ongoing task. As of this moment, 117 SVG flag files have been structured – these can already be used as base flags in the generation. From these flags, 76 already have all their elements with meanings in the semantic dataset and 17 only have some.

10 github.com/lipis/flag-icon-css/
11 netflag.guggenheim.org
12 www.flagid.org
13 www.reddit.com/r/vexillology/comments/2yd77z/

14.3.2 *Generating flags*

As mentioned earlier, there are several ways of producing flags. However, one of our main goals was to be able to maintain the resemblance with the base flag, allowing the identification of the country. For this reason, our system was grounded on two base assumptions: for each flag production, an existing flag would be given as input and the transformations should not go beyond the point in which the original flag is not recognisable anymore – i.e. the produced flags should not be seen as a totally new flag but as a transformation of the original one. This is also motivated by principles of good flag design – "Keep It Simple" (Kaye, 2001) – aiming for small changes and reducing complexity.

At a first stage, the process of producing flags involves the search for elements that match a query word, which are then used to transform the original flag. The search is conducted in three different places: existing flags, a dataset of colour names and emoji.

14.3.2.1 *Existing flags*

We mentioned earlier that structured SVGs could be used as a base flag. However, only flags with associated semantic information can be used to obtain elements to use in the transformation process. This is due to the fact that the search for the input word is conducted using the semantic information – it searches for elements that have the word in their associated meanings. A random selection is then conducted to select a replacement element and a replaced one. Then, the way the blend occurs depends on where the query word is found: if it is in the overall meaning, the full replacement element is used; if it is in the shape meaning, only the shape is used and the colour of the replaced element is applied to it; if it is in the colour meaning, only the colour of the replacement object is applied to the replaced one. All in all, only 522 different words exist in all the collected meanings. This number is not very high when considering that any word can be queried. To increase the chances of successfully finding the query word, we added two other sources of information – emoji and a dataset of colour names.

14.3.2.2 *Colour*

As described in Section 2.3, colour can be used to achieve different perceived meanings when generating symbols to represent a given concept. For example, in the website *Moodjam*[14] the user keeps a record of daily moods using colours.

We use colour to increase the conceptual reach of the system. To produce a colour name dataset, we extended the dataset *color-name-list*[15]

14 `moodjam.com`
15 `github.com/meodai/color-names` (dataset with 18,264 named colours)

by merging it with a list belonging to the *ntc.js*[16] library. In 1308 of the colours, either the hexadecimal value or the name was already on the *color-name-list* list; 215 colours were already on the list with the same name; and only 43 were added to the list (new hex code and nonexistent name). From the resulting colours, we extracted the ones that had names of only one word (e.g. *Tomato* colour), which resulted in a list of 3476 colours. The query word is searched in this list and, if found, the colour is applied to the replaced element.

14.3.2.3 *Emoji*

Emoji are associated with semantic knowledge, as described in Chapter 8. Our position is that this association can be exploited for the visual representation of concepts. This position is the base for the development of the *Emojinating* system (see Chapter 8). For the implementation of *Moody Flags*, we again take advantage of the emoji connection between visual and semantic data.

By having *Emojinating* as inspiration, we decided to add Emoji as a third source of semantic information for the query word to be searched in. To do so, we use the dataset *EmojiNet* – a machine-readable sense inventory with data on 2389 emoji (Wijeratne et al., 2017b) – in combination with emoji SVG images from *Twitter*'s *Twemoji*[17] dataset. When finding emoji that match the word, the system uses them as replacement as follows: if the flag already has a symbol, the symbol is replaced by the emoji; if not, the emoji is added on top of the flag, centred according to a randomly selected element and scaled to fit its bounding box. If the selected element is a triangle, the emoji is scaled a second time for aesthetic purposes.

14.3.3 *"Ever-changing" flags*

The *Net.flag* project is described as an "ever-changing flag of the Internet", which anyone could alter upon visiting the website (Napier, 2002). This concept is related to our approach, questioning the idea of a flag as an object with static nature. The notion of "mutable flag" gains even more significance when combined with a sense of reactivity. We use the term "reactive" (Richardson, 2017) to characterise something that changes according to external input, as defined by Martins et al. (2019c) in the context of Dynamic Visual Identities (DVIs):

> *DVI automatically reacts to external input. A data-driven process is used to autonomously design one or more elements of the VI system. The use of input data (...) enables the VI to become autonomous and alive.*

16 `chir.ag/projects/name-that-color` (dataset with 1566 colours)
17 `github.com/twitter/twemoji`

Reactivity in the context of visual identities is especially relevant for our work as they belong to the visual domain and are related to the concept of "identity" of an entity. One example is the identity of the event *House of Visual Culture* by Edhv,[18] which produced "data minerals" from data related to the event retrieved from the Internet. Another example is the visual identity designed by Neue[19] for the Nordkyn peninsula – the graphic mark was designed to represent the weather conditions at each moment, being updated every five minutes and changing its colour and shape according to the current wind direction and temperature, respectively.

Another example of reactivity can be seen in an interactive installation that employs computational methods in poster design (Rebelo et al., 2019). The system gathers data from the surrounding environment (related to weather and interaction from people) and autonomously generates new compositions with the goal of finding the most efficient way of designing, according to the site where it is placed.

This reactivity to external input can be used to instil a quality of "being alive" into the flags (Martins et al., 2019c). As such, even though any word can be used to produce a flag, our main interest involves producing flags that change according to current events. To achieve this, we follow an approach similar to the one used by Gonçalo Oliveira, Costa, and Pinto (2016), who produce memes using headlines automatically retrieved from the *Google News* RSS feed.

When generating a flag for a given country, the system collects the latest news titles in English that mention the country's name. The second step consists in extracting nouns from the initial news titles by tagging the text using the Javascript Part-of-Speech tagger *jspos*.[20] Then, we analyse the nouns used in all the titles and identify the most predominant ones, excluding the country's name or its abbreviation. After sorting the nouns according to predominance (see topics sorted in Fig. 14.4), the system searches for data to be used in the blending process, as previously described in Section 14.3.2. If no data is found for a noun, the system moves to the next one on the list. This search task is performed until the system finds information (and produces a flag) or until there are no nouns left (no flag is produced).

We developed a web-based interface for the system (described in Section 14.3.5). In each session, information from *Google News* about a country is retrieved only once. This avoids multiple equal queries in case the user tries to produce several flags for the same country. If the user reloads the page, a new query is conducted.

18 www.edhv.nl/
19 neue.no/work/visit-nordkyn/
20 code.google.com/archive/p/jspos/

14.3.4 *Generating explanations*

In most cases, a flag is conceptually grounded – its structure and elements have associated meanings – and changes applied to it should take this into account – e.g. a colour replacement carries a meaning, which will be assigned to the flag. As such, our process of generating a flag consists not only in producing a design but also its explanation. The explanation provides clues of how and why the flag was changed (see examples in Fig. 14.3).

This establishes a link between the visual output and the reasoning behind the production of the flag, thus making evident the conceptual foundation of the generation process.

Explanations follow a predefined structure, which is based on the flag descriptions that we analysed when collecting the data on flag meaning (see Section 14.3.1). The following structure is used:

> [element X] *represents/stands for/symbolises* [Y]

where Y is the query word and X depends on the change nature.

For example, in the case of adding emoji, we defined that X would take the value of "symbol" (see the left side of Fig. 14.2). In contrast, if there was a change of colour, the element X would be composed of the replaced element's name (e.g. "stripe") and the replacement's colour name (e.g. "red"). This posed an issue as, despite the colour name list being useful in finding appropriate colours, it would be confusing for the user to be presented with an explanation such as "The Airforce stripe represents...". To solve this issue, we used Daniel Flueck's extension[21] of the *ntc.js* library, which has a closest colour converter – the colour with the name "Air force" is mapped to the closest standard colour "Blue".

The act of providing explanations is in line with the guidelines for explainability[22] in AI Ethics (Jobin, Ienca, and Vayena, 2019), which is considered to have an important role in design systems – e.g. Zhu et al. (2018) focus on explainability and provide guidelines on how it can be applied in game design. Moreover, the production of explanations can be interpreted as a process of *Framing* as defined by Cook et al. (2019):

> *'Framing' refers to anything (co-)created by software with the purpose of altering an audience or collaborator's perception of a creative work or its creator.*

This process plays an important role in how unbiased observers perceive AI systems and their output. Furthermore, Cook et al. (2019) state that implementing methods for the systems to explain themselves can

21 www.color-blindness.com/color-name-hue/

22 Our interpretation of the principle of explainability and transparency is based on the description from *AI for People*, www.aiforpeople.org/

improve the relationship between user and AI agent. Framing is described as having three aspects: sources of information (e.g. where the meanings of the flags are retrieved from and on what are they based), means of framing (e.g. providing descriptions in natural language for the produced flags) and purposes for framing (i.e. the intended impact on the audience). Llano et al. (2020) address the topic of *Explainable Computational Creativity* and highlight the importance of establishing two-way communication channels between systems and users, allowing the former to explain their decisions in the creative process.

14.3.5 *Interface*

We implemented an interface for the system to allow the user to produce flags according to their preferences (see Fig. 14.2). It consists of two areas: (i) the *configuration area* – where the user defines the parameters for the flag generation – and (ii) the *flag canvas* – where the new flag is shown to the user. The configuration area has two parameters that always need to be provided by the user: the *base country* and *mode of data retrieval*. By default, the automatic mode is selected, and the system uses *Google News* RSS feed to obtain the trending topics to be used in flag production. If the user decides to disable the *auto mode*, the system asks for an extra input: a *topic* to be represented. This way, the user can not only see what the current flag is but also what it would be if a given topic was trending.

14.4 EXPERIMENTATION

In order to assess the perception of generated flags, we conducted a user study. In this section, we describe the study and present the obtained results. Then, in the last part of this section, a general analysis of the system and the generated flags is made.

14.4.1 *Experiment Setup*

We produced a set of five flags (see Fig. 14.3): two resulted from *colour replacement*, two from *symbol replacement with emoji* and one from *emoji addition*. These flags were automatically generated using the news at the moment of generation and selected by the authors. The participants were informed that they would be presented with flags computationally generated using real-time news. They were also asked not to search for any information while conducting the experiment nor change any answers.

For each flag, the participant was asked to answer questions from two different sections. In the first section of the survey, the only given information about the flag was the generation day and the users were

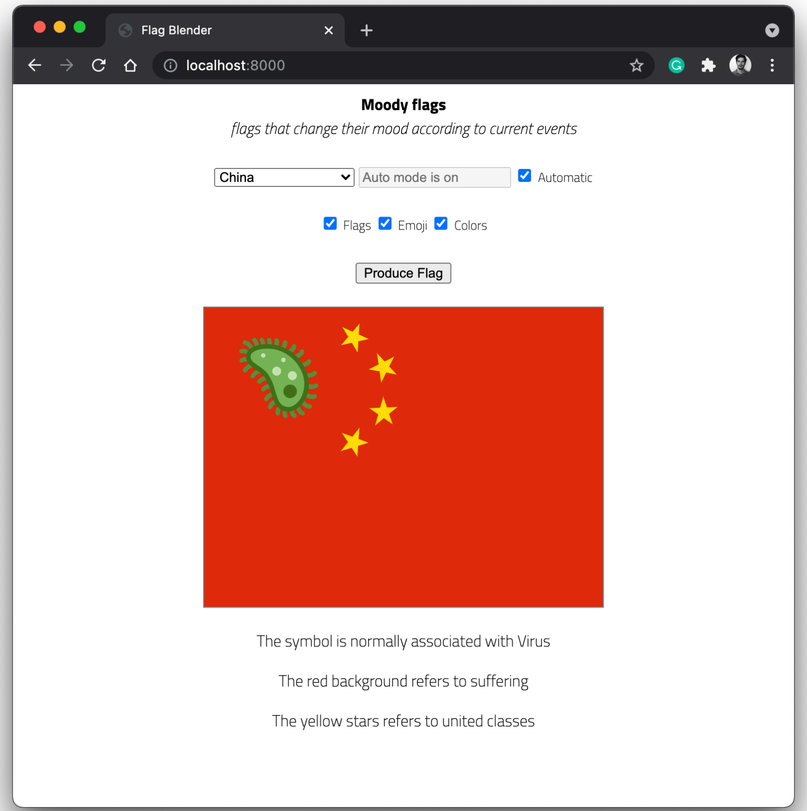


Figure 14.2: A web-based application showing a generated flag based on China's flag and its explanation (produced on February 11^{th}, 2020), in which the first sentence corresponds to the changed element (note: the sizes were intentionally changed to increase the legibility of the figure).

asked to answer or complete the following open-ended questions or tasks:

Q1 If you know which country is represented in the flag, please write the name.

Q2 There is a change in the flag. Describe what you think the change was.

Q3 What do you think that the change represents?

In the second section, users were told which country was represented, what the change was and the topic on which the change was based (e.g. flag of Brazil; the background colour was changed into dark grey using news about the oil spill). Then, the user was asked to answer the question below with a number between 1 (very bad) and 5 (very good):

Q4 Is the flag a good representation of the news?

Figure 14.3: Flags used in user survey and produced explanations, automatically generated on 15$^{\text{th}}$ November 2019.

14.4.2 *Results*

The survey was conducted with 16 participants, with ages between 26-44. The results obtained can be seen in Table 14.1. For Q1, we considered correct answers the ones that referred the country of the base flag. Also, in flag #4 we considered answers such as "Argentina + Brasil" as correct due to the fact that the blended flag has both flags. On the other hand, in flag #3 we considered answers such as "United Kingdom" as wrong (despite the United Kingdom flag, or *Union Jack*, being included in the Australian flag) as the participant is clearly not familiar with the Australian flag. For Q3, we considered correct answers the ones that referred the word used to generate the flag. However, in the case of flag #5, we considered correct three answers from Q3 that did not mention "spy", as the participant had mentioned it in Q2 – e.g. Q2 "change of the icon into a spy figure", Q3 "leak of information". In fact, one of the participants commented that they had answered to Q2 that the change was the addition of a spy icon but on Q3 they had thought that "spy" would be too simple.

When observing the results, one of the things that stand out is that for all the flags, except for #5, the majority of the participants could identify the country and the change that occurred – indicating familiarity with the original flags. It is interesting to see that, in the case of flag #5 (Lithuania), despite the participants being unfamiliar with the flag, they could identify both the change and the meaning – which reflects the perception advantages of emoji. On the other hand, in flag #1 no one could identify the meaning, despite everyone knowing the original flag (Brazil). When analysing the answers by the participants to Q3 of flag #1, four out of the 16 mentioned the burnt Amazon forest, which was a highly discussed topic at the time and a possible interpretation of changing the green to dark grey. Similarly, in flag #2, in which 11 out of 16 people got the answer right to Q3 ("elections"), two other people gave an answer related to political instability and another one gave an answer related to a referendum – both answers, despite not matching "elections", are aligned with the replacement of the symbol by a voting poll seen in the generated flag and with the situation of the country at the time. It is also worth mentioning that some of the participants that

Table 14.1: User study results of each of the generated flags for questions Q1 (country), Q2 (change), Q3 (meaning) and Q4. The results of Q1-3 are expressed in percentage of right answers and Q4 is expressed using the interval from 1 (very bad) to 5 (very good).

				right answers (%)			Q4 (1-5)	
#	**original**	**word**	**change**	**Q1**	**Q2**	**Q3**	mode	median
1	*Brazil*	*Oil*	background colour	100.0	100.0	0.0	4	4
2	*Spain*	*Election*	symbol replacement	100.0	100.0	68.8	5	5
3	*Australia*	*Fire*	background colour	68.8	87.5	25.0	4	4
4	*Argentina*	*Brazil*	symbol replacement	93.8	93.8	31.3	3	3
5	*Lithuania*	*Spy*	emoji addition	25.0	43.8	31.3	5	4.5

did not know the meaning of flag #3, which had a background colour change into red, submitted answers that could somehow be linked to that colour, for example "blood", "massacre" or "terrorist attempt".

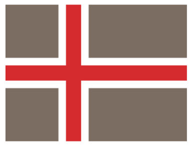
Regarding quality, four out of the five flags obtained a quality of topic representation of good or very good by most participants. The results also seem to reflect the easiness of understanding emoji (see flags #2 and #5). However, flag #4 also uses emoji and had the lowest results. We cannot be certain, but we believe that this was due to how section 2 of the survey was designed for this flag. The user was presented with an explanation giving especial focus to football – "The symbol was changed using news about the football match between Argentina and Brazil" – but that meaning was not reflected on the blended flag.

14.4.3 *Discussion*

The capacity of our system to generate flags is highly dependent on the existence of semantic knowledge, which is used to find possible changes to be made. We believe that by adding three sources of semantic information (meanings of existing flags, emoji semantic data and colour names), we have increased the likelihood of success. However, it is impossible to guarantee the production of good results. For example, one case in which the system has few results is the word "state": in terms of data on existing flags, the only matches are star-shaped elements (e.g. the white stars in the United States flag); by considering emoji data, the system is able to find 255 different emoji, most of which are flags themselves; and using colour names, there is no match for "state". Two flags produced for Iceland are another example. The resulting flag changes depending on which data is available (Fig. 14.4): if the system only uses data of existing flags, it is not able to produce any blend; if it uses emoji data, it is able to find information for the third trending topic ("Christmas" represented using a Christmas tree); and if it uses colour names data, it can only find information regarding the

1. **"Icelandic"**, count = 16
2. **"Airwaves"**, count = 5
3. **"Christmas"**, count = 4
4. **"Icelanders"** count = 4
5. **"Bribery"**, count = 3
6. **"Namibia"**, count = 3:
   e.g "Icelandic Operation in Namibia under Scrutiny"

Iceland
"Christmas"



Iceland
"Namibia"

Figure 14.4: Flags generated for Iceland using different semantic data sources (emoji and colour).

22:47 UTC

1. **"Argentina"**, count = 8
   e.g "The Return Of King Messi: What we Learned From Brazil Versus Argentina"
2. **"Oil"**, count = 8
3. **"China"**, count = 5
4. **"Reform"** count = 5
5. **"Bolsonaro"**, count = 5
6. **"Amazon"**, count = 4

...

Brazil
"Argentina"
22:47 UTC

22:53 UTC

1. **"Oil"**, count = 8
2. **"Argentina"**, count = 7
3. **"Bolsonaro"**, count = 6
4. **"Amazon"**, count = 5
5. **"Spill"**, count = 5
6. **"China"**, count =4

...

Brazil
"Oil"
22:53 UTC

Figure 14.5: Mood shift due to football match Brazil vs Argentina, on 15th November 2019.

6th trending topic "Namibia", which matches the name of one of the colours in the dataset and is also the name of a country.

From the conducted user study, it is clear that the meaning of the changes is not easy to guess and is very dependent on the user knowledge about the corresponding country and its current situation – only one of the flags had a correct response rate to Q3 (meaning) above 1/3. This leads us to conclude that the changes in the flag should have more impact within the corresponding country than internationally – as stated by Matusitz (2007) "vexillological symbols are displayed to the whole world, but are only understood by like-minded individuals", which is in accordance with findings of difficulty in flag identification (Morales-Ramirez, 2018). For this reason, further studies with citizens of each country would be needed to fully analyse the impact of the flags – none of the participants was a citizen of any of the countries with changed flags.

One interesting aspect of the project is the ability to observe this "ever-changing" identity or, using the term that we adopted, the *mood changes* of the country. An example of mood changing was observed on the 15th of November 2019, due to a football match between Brazil and Argentina (see Fig. 14.5). During the hours before the match, the flag of Brazil was always retrieving "oil" as mood from the oil spill. Then, Messi scored and the mood changed, resulting in a different flag – for roughly five hours the mood stayed with "Argentina". Six hours later, it alternated between "Argentina" and "oil", and later on it went fully back to "oil".

Despite being different flags (oil-driven and Argentina-driven), it is possible to identify the resemblance with the original Brazilian flag.

Figure 14.6: Examples of flags generated on 15th November 2019. Below each flag, the country of the original flag and the trending topic used in the generation are identified.

This aspect was of particular importance to us and the reason why we chose to only apply one change and avoid adding many elements, which would increase the complexity of the flag. Nonetheless, it would be interesting to see different trends affecting the flag at the same time, choosing the element to change according to its salience (i.e. impact on the overall aspect of the flag) to match the trendiness degree – the more trending the more salient the changed element should be.

Even though the system only makes a change, some flags have few characteristic elements, which may end up being replaced and, consequently, the connection to the original flag is lost – an example is the flag of Saudi Arabia in which the symbol (an Arabic inscription and a sword) is replaced by a bird to symbolise *Twitter* (Fig. 14.6). Therefore, the applied changes, despite being simple, can go from *subtle* – unidentifiable for most people – to *disruptive* – possibly triggering a sense of discomfort on the viewer, who might see familiar elements but no longer relate the flag to their country, creating a gap on the notion of identity. This aspect gains even more importance if we consider that the citizens of a country may have different opinions regarding the national flag (Satherley, Osborne, and Sibley, 2019; Wright, 2011).

It is also possible to observe the effect of the same topic on different flags, for example “oil” in Fig. 14.6. As we have not implemented a system to deal with differences in salience, the visual change is similar, for example in the flags of Brazil and Norway, even though the seriousness of the news varies in degree – in the Brazilian one, it should look more catastrophic due to the gravity of the situation. A similar effect occurs in the blend using the Pakistan flag, which is based on the topic "children" and results in a blend that applies a green colour to the symbols of the flag. Despite using the green colour, which is normally associated with good, the news behind the trending topic are far from positive (e.g. "An HIV Crisis Among Pakistan Children").

Moreover, some changes might make more sense when applied to certain elements. For example, Angola was also getting the "oil" trending topic and could have it applied to its cogwheel, which is associated with industry. This would make sense if we look at some of the news, e.g. "Angola oil production falls in October to 1,356 million barrels per day". Another example can be observed in two flags generated for Brazil using "Oil": in Fig. 14.3 the dark grey was applied to the green background; and in Fig. 14.5 it was applied to the blue circle. The latter version would be more suitable as the oil spill occurred in the (blue) sea, whereas the former version can be more easily mistaken for another topic: the Amazon fires. As such, a future development might involve taking into consideration the meaning or characteristics of the replaced element – "burnt" being applied to the green of Brazil flag or using Angola's cogwheel to represent industry-related topics.

Incorrect behaviours of the system also occur. For example, when producing flags for Jordan it retrieves incorrectly matched news, getting news about Michael Jordan, instead of the country, leading to the trending topic "Basketball" and resulting in the orange colour being used (Fig. 14.6). Similarly, when using the topic "Trump", the system obtains a musical instrument emoji instead of something that represents Donald Trump (President of the United States at the time). It is also important to mention how using elements from other flags might have a different effect than expected. Some of the elements and associations are culture-specific (Becker et al., 2017; Morales-Ramirez, 2018) and might not have the same interpretation in all countries.

Another subject concerns the production of lower quality flags. This can happen by placing an element in an unsuitable place (e.g. the triangle in two of the flags in Fig. 14.7 looks like a play icon) or using a given colour and making the flag look like an existing one, losing its initial identity (e.g. the first flag in Fig. 14.7 resembles the Russian one).

Despite the existing issues, the results of the user study showed that, even if the user does not know the flag, it might be possible to infer some meaning. This can be exploited by using the flag to call the attention of the user to countries in which something relevant is happening. One example of this was identified in the study: none of the participants was able to link flag #1 (Brazil) to the huge oil spill that had occurred. As such, it could be possible to use the flags as a way of raising awareness, similar to what was done in *Double Standards* (Pater, 2012).

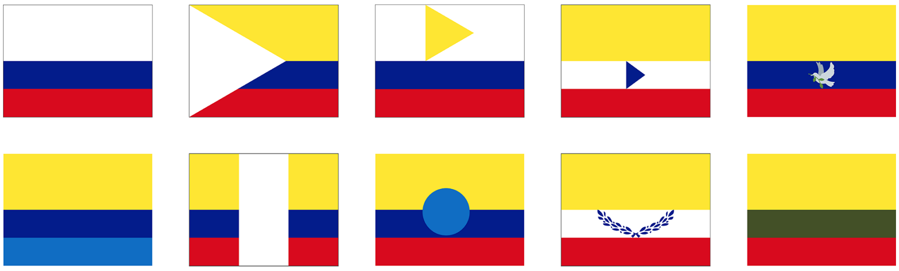


Figure 14.7: Flags generated from the flag of Colombia using the topic "peace".

14.5 ETHICAL CONSIDERATIONS AND POTENTIAL IMPACT

With *Moody Flags*, our goals go beyond the generation of flags. Part of our motivation for developing this system is to have an impact on the observer, as opposed to a generation for mere aesthetic or representational purposes. In the following section, we provide more detail on this subject.

The limits of the use of a national flag have been a topic of debate. As we have seen, flags are prone to be misappropriated – aspect highlighted in the *Double Standards* project (Pater, 2012). Moreover, as symbols of a nation, they are often used in acts motivated by political reasons – e.g. flags being burnt in protests. For these reasons, several cases exist of controversy around what is considered legal and what is to be seen as flag desecration (Goldstein, 2019; Marinthe et al., 2019). However, the limits are often blurry and lead to strong yet opposing reactions when they are tested. One example is the installation *What is the Proper Way to Display a U.S. Flag?* by Dread Scott,[23] which showed two images featuring the American Flag, one of which displayed a flag being burnt, and encouraged the audience to write responses to the question in the installation's title. Upon writing a response, the audience had the option of standing on the flag. The installation triggered very strong reactions – from thank you messages to death threats. But more importantly, led to a discussion on what is a misuse of the flag and the legality of such. It is clear that there is a significant difference between purposely destroying a flag and using it to communicate an idea, with the latter being especially important for artistic purposes (Hartvigsen, 2018). Focusing on what we are proposing in this paper, to what extent do flags actually represent constantly evolving nations when they are subject to rules often against change and transformation? In addition, people are not always receptive to changes in the national flag, as it deals with questions of their own identity (Osborne et al., 2016). This immutability reaches the point that the flag design stays the same but the meanings change – e.g. the colours of the Por-

23 www.dreadscott.net/

Figure 14.8: Flags generated on November 15th 2019 and July 15th 2020. Below each flag, the country of the original flag and the topic used in the generation are identified.

tuguese flag went from a political connotation (party colours) to more general ones (e.g. with green being associated with "hope").

We intend to contribute to this discussion by questioning the unchangeable status of a flag. As such, we identify several topics that we believe our system has the potential to have an impact on:

- *Own sense of identity:* the feelings towards a flag vary from person to person: some might not have a strong connection to this symbol; others might proudly display it on the window to convey a sense of national support (e.g. in some countries flags are often hanged from windows in support of the national football team); and, possibly, there may be citizens that feel misrepresented by the flag (Wright, 2011). In any case, we believe that changing a country's flag will lead to a sense of "discomfort" by creating a gap between the original symbol and an altered version, possibly making people wonder if they still identify themselves with it;

- *Evolution of daily topics:* flags are objects that often have a very slow evolution – they stay the same for long periods. Our system brings changes in this regard by allowing flags to adapt to current events (see flags produced in two different dates in Fig. 14.8). Such approach enables the user to observe a constant change in the flag, a consequence of changes in the "mood" of the country;

- *Event highlighting:* despite living on what can be called a "global village" (McLuhan, 1962), there are many events that often go unnoticed, even though they deserve our utmost attention – the case of a huge oil spill that was not widely known, identified in the user study. Our system has the potential of being exploited as a visualisation tool with the goal of highlighting such events. Using flags to call the public attention has been explored in the past, e.g. in the *Double Standards* project (Pater, 2012).

14.6 SUMMARY

The flag of a nation serves, among many things, to build and maintain the sense of national identity, representing the country, its people and its history. In this chapter, we propose a different use for a flag – the representation of a country's mood at each moment. We presented a system that produces variations of national flags according to news titles retrieved from the *Google News* RSS feed, by using semantic information from different sources.

The system relies on two base assumptions: (i) when generating a flag, an existing flag would be given as input and (ii) the changes made should allow the initial flag to be recognised. The produced flag should be perceived as a transformation to the original one, thus allowing the observer to identify the country.

The first step of the flag production process consists in searching elements that match a query word. These are then used to change the initial flag. The element search is conducted in three places:

- *Existing Flags:* we produced a flag dataset that included both visual data and semantic data. Using this dataset, a search for the input word is conducted on the meanings assigned to elements of existing flags;
- *Colour Names:* we merged existing datasets to produce a list of 3476 colours and associated names (e.g. `#ef4026` has the name "Tomato"). This list is used to search for the input word;
- *Emoji:* we use the *EmojiNet* (Wijeratne et al., 2017b) dataset to find emoji based on the input word, similar to what is done in the *Emojinating* system.

The transformation made to the flag depends on the type of elements found: if the input word is found on a colour name, the colour is applied to an element of the base flag; if the element found is an emoji, it can either be added to the flag or replace an existing symbol. In addition to producing a flag, the system also presents the user with an explanation for the changed or added elements.

In order to assess the perception of generated flags, we conducted a user study with 16 participants. The results show that the participants can identify the original flag but they have certain difficulty in identifying the meaning of the changes applied to the flags. We also highlighted the potential impact of generated flags, which goes from raising awareness (to a certain event) to creating a sense of awkwardness by affecting the notion of identity.

Part V

TOWARDS VISUAL CONCEPTUAL BLENDING

Visual Blending can be used for the production of visual representations of concepts. However, Visual Blending on its own does not necessarily have a conceptual grounding and may lead to nonsense solutions. To address this issue, a Visual Conceptual Blending approach, in which Conceptual Blending and Visual Blending are combined, may be used to improve the quality of the solutions of concept visual representation. In this part of the thesis, we address the topic of Visual Conceptual Blending, present a roadmap for its implementation and describe useful resources and approaches.

15 FROM CONCEPTUAL BLENDING TO VISUAL BLENDING AND BEYOND

At the Dagstuhl Seminar 19172 *Computational Creativity Meets Digital Literary Studies* took place a talk titled *From Conceptual Blending to Visual Blending And Back* by Cunha and Cardoso (2019), in which the importance of having a conceptual ground for producing visual blends was highlighted. In this talk, Conceptual Blending and Visual Blending were first introduced and then the connection between the two was addressed, emphasising the potential of their combination as what can be referred to as Visual Conceptual Blending. By following a Visual Conceptual Blending approach, visual blends can be produced based on and guided by conceptual reasoning. As such, a visual conceptual blend can be seen as a visual blend conceptually grounded and complemented by a conceptual layer developed through elaboration.

In this chapter, we outline a roadmap for visual conceptual blending, oriented towards its implementation as a computational system.

This chapter is based on the work described in the papers by Cunha, Martins, and Machado (2020d).

15.1 CONTEXT

On the topic of blending, existing computational systems can be placed on a spectrum that has Conceptual Blending on one end and Visual Blending on the other end. In Chapter 4 we introduced the reader to both Conceptual Blending and Visual Blending, describing the different aspects that they encompass and also the existing research related to them. Below we summarise the main features of Conceptual Blending and Visual Blending.

Conceptual Blending:

- *input*: two or more domains (mental spaces);
- *output*: blend domain, which has a partial structure from the input domains but also an emergent structure of its own;
- *mechanisms*: composition, completion and elaboration.

Visual Blending:

- *input*: two or more objects (images);
- *output*: an image showing an integration of the two input objects (the visual blend), which are still recognisable and enable the

viewer to infer possible associations between the concepts that the input objects may represent;

- *mechanisms*: juxtaposition, replacement and fusion (simplified structural taxonomy).

On the Conceptual Blending end, approaches focus on the conceptual level and the generation of visual output is done for visualisation purposes (e.g. Pereira and Cardoso, 2002). On the Visual Blending end, approaches focus on the visual level and mostly consist in the combination of input images (e.g. Correia et al., 2016). In the middle of the spectrum is what can be referred to as Visual Conceptual Blending.

Most work that addresses Visual Blending falls short when it comes to the conceptual level, often relying on the user to establish a connection between the visual and the conceptual levels (e.g. Chilton, Petridis, and Agrawala, 2019) or simply using a direct mapping between the input concepts and the visual representations used in the blend (e.g, Zhao et al., 2020).

With the work described in this thesis, we aimed at improving the connection between the conceptual level and the visual one. In the *Pig, Angel and Cactus* experiment,[1] described in Chapter 6, we introduced the idea of a hybrid blending process, bringing the conceptual and the visual levels together. With *Emojinating*,[2] described in Chapter 8, we highlight the combination of visual and conceptual blending as visual conceptual blending and we introduce a mechanism for *concept extension*, increasing the role of conceptualisation.

Visual Conceptual Blending has also been mentioned by other authors. Karimi et al. (2018b) present their work as a computational model for generating visual conceptual blends in the domain of sketching. However, the core of the model is more related to conceptual shifts – retrieving sketches similar to an initial one – than with visual blending, which is later presented as a possible application and not intended as an automatic process.

Chen (2019) introduces their work as a generative model for visual conceptual blending. As described in Section 3.2, the work consists of a Generative Adversarial Network (GAN) model capable of generating images that depict a blend between two different concepts (e.g. a *spoon* and a *leaf*). In a way, this approach is still distant from visual conceptual blending, as it lacks in terms of conceptual level.

More recently, Ge and Parikh (2021) present their work as an approach to the generation visual conceptual blends through the use of a language model to identify objects to blend and a text-based image generation model to produce the visual blends. The work is aligned with the idea of visual conceptual blending that we defend in this thesis,

1 Authored by Cunha et al. (2017).
2 Authored by Cunha, Martins, and Machado (2018b)

Figure 15.1: Animal visual blends. All blends were created by Arne Olav (`gyyporama.com`), with the exception of "elephaneleon".

facilitating a connection between the conceptual and the visual level. More detail on this work has been given in Section 3.2.

Despite providing valuable clues on the direction towards a possible model on visual conceptual blending, these systems cannot be considered as one. In our opinion, they fail to address several topics that we believe are important when building visual conceptual blends. Moreover, a concrete formulation of what visual conceptual blending may involve is still lacking. In this chapter, we will focus on the topic and attempt to contribute to a possible formulation.

15.2 ANALYSIS TO VISUAL BLENDS

According to Pollak et al. (2015), there are still many open questions regarding the production of blends. By investigating human creations and identifying patterns, it is possible to address these questions and possibly find a direction for the blending process, eventually allowing the automated generation of blends (Pollak et al., 2015). Joy, F. Sherry Jr., and Deschenes (2009) conducted an analysis of blends based on human perception by analysing conceptual blending in advertising. Bolognesi, Heerik, and Berg (2018) built a corpus of visual metaphors that have been analysed and annotated on different dimensions of meaning. Petridis and Chilton (2019) focus on how people interpret visual metaphors and identify causes for misinterpretation.

For our work, the most interesting example of blend analysis was conducted by Martins et al. (2015), who conducted an online-survey questionnaire in which participants were asked to evaluate criteria assumed to be related to the quality of blends. Martins et al. (2015) used visual blends between two animals (see Fig. 15.1) and tried to identify what humans perceive as a good blend. These blends used *fusion* and were focused on *perceptual features*, e.g. *colour*, *texture* or *pattern*.

Upon analysing the blends (Fig. 15.1), one observes that colour cannot be considered the main reason for conducting the blend – i.e. animals are not blended on the basis of similar colour – but as a way to produce a good blend by achieving a fully integrated blend. Nonetheless, in some blends colour alignment of the input animals seems occur (e.g. *pengwhale* or *guinea lion*). In the same way, proportion is also not the ground for blending, as several examples exist of strange proportion between head and body (e.g. *snorse*). It leads to the conclusion that the selection of the input animals was conducted without any apparent reason or conceptual grounding. Regarding the mapping that leads to the blend, one can see that it is mostly based on element category similarity (e.g. head of the snake is mapped to the head of the horse). Nonetheless, Martins et al. (2015) give special attention to elaboration: *name building* and *context creation*.

Another example of blend analysis is described by Chilton, Petridis, and Agrawala (2019), who stated that they observed blend examples and tested theories to come up with a design pattern – they identified shape as a particularly important feature in visual blending. Based on this, they developed a workflow for producing visual blendings based on an abstract structure: blend two objects that have the same basic shape but other identifying visual features. This example contrasts with the one from (Martins et al., 2015), as they use a completely different feature. In addition, whereas Martins et al. (2015) only used blends of animals (*fusion* blend type), Chilton, Petridis, and Agrawala (2019) analysed visual blends of objects based on *replacing fusion type* – see the taxonomy that we proposed in Section 5.2, based on the work by Peterson (2018).

15.3 ROADMAP FOR VISUAL CONCEPTUAL BLENDING

In this chapter, we aim to outline a model for the production of visual blendings with a strong conceptual grounding. In a process of visual conceptual blending, despite the output being a visual blend, it does not merely consist in the task of producing a merge of two initial visual representations. Instead, the core of the process has to do with conceptual reasoning, which serves as a base for the actual process of visual blending. This contrasts with the description of the constituents of a visual blend.

In visual conceptual blending, the focus is not the transformational task of mixing two images but the whole process of producing visual blends that are based on a conceptual reasoning and present themselves as a result of a knowledge-based process. In fact, from our perspective, the output of a process of visual conceptual blending is not only an image but also a set of conceptual elaborations. A visual conceptual blending has context, it is grounded on a justification that in-

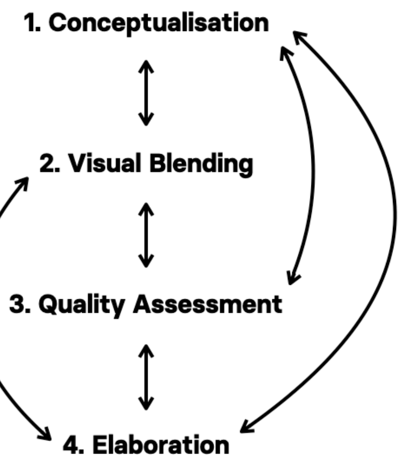


Figure 15.2: Stages of Visual Conceptual Blending

dicates the relevance of the blend. It can also be given a name that may not even be related to the original concept.

In this section, we describe a model for the production of visual conceptual blends. Our main goal is to provide a roadmap rather than a final blueprint, providing a broad description that mentions all the topics that we deem important to build such a model. Our roadmap is composed of four main stages (Fig. 15.2): (i) Conceptualisation; (ii) Visual Blending; (iii) Quality Assessment; and (iv) Elaboration.

Despite presenting it as a sequence of stages, the reader should understand that their order may vary, not being fixed and allowing stages to be revisited if needed (e.g. one may need to return to *Conceptualisation* after the *Visual Blending* stage).

15.3.1 *Conceptualisation*

By just focusing on the visual blending between two objects, we can generate a large (if not infinite) number of possibilities (see examples in Fig. 15.3). However, these are not guaranteed to be grounded on knowledge. For example, for the visual blend between a snake and a horse, instead of the logical category-category mapping between heads (seen in Fig. 15.1), it would be possible to produce a blend in which the head of the snake replaces the tail of the horse (Fig. 15.3). Such a blend would have a very low conceptual grounding as no apparent logical mapping was performed. Conceptualisation is what distinguishes mere generation from something with a strong conceptual grounding (resultant from a process of reflection) and consequently visual blending from visual conceptual blending.

Conceptualisation can occur in at least two situations: selection of input concepts and mapping between previously given ones. Most examples described in this chapter fall under the latter case – the input

Figure 15.3: Visual Blends between horse and snake.

concepts to use in the blend are previously chosen. Nevertheless, in a general model, an initial step may concern the identification of potential candidates for a visual conceptual blending (Gonçalves et al., 2015). The topics addressed here can be applied to these two scenarios.

In fact, the process of conceptualisation may lead to the retrieval of related concepts for the production of the visual blend, as is the case of our approach with *Emojinating* (Chapter 8). In such a case, there may not be a direct relation between the original concept and the visual blend. This is especially evident when the original concept consists of one word – it is necessary some sort of expansion to provide a foundation for a process of visual blending to occur. In these cases, the original concept can be visually represented by resorting to related concepts, e.g. *freedom* represented using a related concept "universal right" (Fig. 15.4). Moreover, the interpretation of the resultant visual blend can lead to a third concept, for example "travel the world". Therefore, the levels of conceptualisation of a visual blend can vary. In fact, the process of conceptualisation can reach high degrees of complexity – e.g. using a process of conceptual blending based on structural alignment techniques to produce analogies from structures such as mental spaces, as was the case with the *Blender* and the *Pig, Angel and Cactus* experiment (Chapter 6).

Figure 15.4: Freedom blend by *Emojinating*

On the other hand, a process of visual conceptual blending can have different motivations and therefore different goals. For example, the process can be used for concept representation, in which case a literal representation may be preferred. Another possibility is the production of visual metaphors, in which case the goal will be more creative.

In the end, the conceptualisation stage consists in answering the question: *what is behind the blend?* How this question is approached depends on the starting point. For example, if we already have a double-word concept it is more related to how we blend the two concepts – finding a justification for a blend. If the starting point is a single-word concept, we face a somehow open search for potential blends – which is good if we have enough knowledge. A search for potential blends using single-word concepts is conducted, for example, in *Emojinating* and in the work by Ge and Parikh (2021).

The issue here is also related to one of the problematics of conceptual blending: finding a common ground between the input concepts to allow the blending to occur (Eppe et al., 2018). In general, several characteristics can motivate the process of blending, e.g. *conceptual features* (e.g. name or affordances) or *perceptual features* (e.g. shape or colour).

Perceptual features: one way of grounding the blend is by using perceptual features (low-level features), e.g. shape or colour. The usefulness in perceptual features is especially relevant when these include prototypical elements – i.e. what most identifies a given concept, e.g. the *nose* and the *tail* in a *pig*. An example is the work by Karimi et al. (2018b), in which blend possibilities are found using a process of conceptual shift based on shape comparison.

Obviously, these characteristics are very dependable on the representation used – e.g. if only images in black and white are to be used (as is the case with the *Pig, Angel and Cactus* experiment), colour loses relevance. The mappings based on perceptual features always depend on the situation.

Affordances: another way of finding blend possibilities is related to affordances and their modelling using, for example, image schemas. Such may help in guiding how the visual blending should be conducted – e.g. using the schema CONTAINMENT with icons of *money* and *building* to represent *bank*. We will return to this topic in Chapter 16, in which we address how visual representations can be produced through visual blending using image schemas.

Naming: a third possibility is related to the name – e.g. finding homophones such as "waste of money" and "waist of money", or "racket" and "rawcket" (see Fig. 15.5). As an example, Veale and Al-Najjar (2016) explore the invention of colour names.

Figure 15.5: Rawcket by *les.creatonautes*

15.3.2 *Visual Blending*

Existing visual blending systems can be divided into two groups based on the type of rendering: *photorealistic* and *non-photorealistic*. These two types have great differences in terms of how the visual blending process occurs. A photorealistic visual blending may require *computer vision* and *image processing techniques*, whereas a non-photorealistic visual blending that uses Scalable Vector Graphics (SVGs) is easier to conduct (as we have argued in our experiments, for example with *Emojinating* in Chapter 8).

In either case, a process of visual blending involves two main decisions: *which objects to combine* and *how to combine them*.

15.3.2.1 *Connection between Conceptual and Visual*

Most of the visual blending examples that are grounded on a process of conceptual blending consist in a simple visualisation of the blend (e.g.

Figure 15.6: Dinosaur blends with *salad* (left), *fish* (middle) and *park* (right). Source: left and middle by *les.creatonautes* (instagram); right by the author.

Pereira and Cardoso, 2002). An exception can be seen in the *Pig, Angel and Cactus* experiment, in which two types of network structures were used: one corresponding to the mental spaces of the input concepts and another corresponding to the visual structure of the visual representation (see Chapter 6). The two types of structure were aligned to produce visual conceptual blends. In this case, the system is considered a hybrid blender, as the blending process starts at the conceptual level and ends at the visual one. However, this situation is uncommon, as in most cases it is not possible to align the conceptual layer with the visual one – such data would have to be manually built. One possibility would rely on an analysis of the images to produce a network structure (structure extraction). With SVGs this would be easier to implement than with raster images, which would require the use of techniques such as concept detection (Zhou, Jagadeesh, and Piramuthu, 2015).

In any case, in the same way, the visual level is based on what is produced on the conceptual level, the conceptual level also needs to take into account the character and features of the used representations.

In addition to the simple exchange of parts, there are several aspects that have to be taken into consideration. For example, questions related to semiotics are especially important (as described in Section 2.3), as colour, shape and other visual features can easily affect meaning.

15.3.2.2 *Type of Blend*

Another important topic concerns the mechanisms used to produce the blend, as different types of blends may be more suitable to specific types of concepts and visual representations. This way, the choice of which blend type to use should take certain aspects into consideration. First, it should consider the relationship between the categories of the concepts being blended. For example, it is completely different to blend *dinosaur* with *park* or *dinosaur* with *fish* (see Fig. 15.6). The former case involves an *animal* and a *location*, which makes it more suitable to have a *juxtaposition*. In the latter case, both concepts are *animals* and, as such, a *fusion* might be more appropriate. In Chapter 17, we will return to this topic by using an emoji categorisation to study the role of different categories in visual blends.

Then, since the process of blending involves visual representations (e.g. *icons*), the appropriateness of the blend type also varies depending on the type of representation being used. For example, in the "dinosaur" "fish" the animals are very different and that will have an impact on how the blend is conducted. Moreover, it is completely different to blend two representations that show the full body of the animal and one that shows a full body and another that only shows the head. For the blends shown in Fig. 15.1, it is likely that the author had to select the images that best matched one another.

A possible approach may resort to the notion of integration networks, which was briefly addressed in Section 4.1.2. In conceptual blending, the cross-space mapping between input spaces may use different types of integration networks. These depend on the frames that exist in the input spaces – e.g. in a Mirror network, both inputs contain the same frame. A similar perspective can be explored in terms of the structure of input visual representations (what sort of structures exist in the visual representations and how may these affect the blending process), which may lead to an eventual connection between the conceptual level and the visual one.

15.3.3 *Quality Assessment*

When producing blends, it is crucial to have a measure of quality to identify good solutions. In certain situations, the blend production can be considered an open-ended problem, in which case including the user in the cycle may be advantageous. Nonetheless, several types of quality assessment exist – some may be more suitable for certain goals than others.

In fact, Martins et al. (2015) pose several questions regarding quality assessment: "How 'semantically far' should the input spaces be to produce a good blend?", "Is there a correlation between the quality of blends and the number of elements for projection?" or even "Are all the optimality principles required to produce good blends?" In this section, we present some types of quality measures that can be used to assess how good a blend might be.

15.3.3.1 *Argumentation*

Confalonieri et al. (2015) propose the use of argumentation to evaluate and iteratively refine the quality of blended computer icons. The authors introduce a semiotic system, which is based on the idea that signs can be combined to convey multiple intended meanings. Despite this, no evidence of a possible implementation was provided.

15.3.3.2 *Optimality Principles*

Figure 15.7: Logos of *Ubuntu*-based distributions: *Kbuntu* (top) and *Xbuntu* (bottom). Adapted from: (Kowalewski, 2008)

Fauconnier and Turner (1998) propose a list of optimality principles that can guide the process of conceptual blending. These principles are not trivial to computationally model and are normally used at the conceptual level. Nonetheless, it is also possible to use them to validate the blend on the visual level, as Kowalewski (2008) demonstrate by analysing the formation of logos and product names in terms of usage of optimality principles.

Even though these principles are considered responsible for generating consistent blends (Martins et al., 2015), they should not be regarded as "rigid laws" but as flexible guidelines (Kowalewski, 2008). We provide a description of these principles below:

- **Integration:** the blend must constitute a tightly integrated scene that can be manipulated as a unit. It should be a coherent, self-contained and unified structure (recognised as a whole). Martins et al. (2015) identify Integration as the most important principle. According to the definition of visual blend as an integration of input images, this principle should be fulfilled by nature.

- **Topology:** the elements projected into the blend should maintain the same neighbourhood relations as in the input space. Even though Martins et al. (2015) indicate that topology is not relevant, according to Kowalewski (2008) it can be useful for example in terms of spatial organisation by placing elements in the blend according to the configuration of one of the input visual representations (e.g. maintaining the existence of a central element, laying new elements according to the centre-periphery scheme, as the mouse element in *Xbuntu* logo in Fig. 15.7). This principle was given particular importance in the *Pig, Angel and Cactus* experiment (Chapter 6) due to the way visual relations in the input were used to map and place elements.

- **Web:** the blend as a unit must maintain the web of appropriate connections to the input spaces, so that an event in one of the input spaces implies a corresponding event in the blend.

- **Unpacking:** this principle consists in the easiness of reconstructing the inputs and the network of connections from the blend. The input concepts should be recognisable from the elements of the blend, through the identification of the input visual representations or parts of them. Figure 15.7 shows an example of this, in which the *Kbuntu* logo uses the *circular structure* of Ubuntu's and the *cogwheel* element from KDE's. This is related to the use of *prototypical parts* in the blend, which we explored in the *Pig, Angel and Cactus* experiment (Chapter 6).

- **Relevance (or Good Reason):** if an element appears in the blend it should have some kind of significance / meaning. This is easy to observe in the application of colour to a blend should be based on the input visual representations – e.g. using green from a *snake* in *snorse*.

Two other principles are *Intensifying Vital Relations* and *Maximising Vital Relations*. However, in this context, we could not provide a clear usefulness for them. In addition to being sometimes vague and difficult to implement, not all the principles are compatible with each other (Martins et al., 2015). Moreover, choosing some over others may lead to a variation in the creativity degree of the blends (Martins et al., 2016).

15.3.3.3 *Visual Analysis*

Assessing quality can also concern visual aspects. Two examples are: *overall complexity* and *area exchanged*, both explored in the co-creative approach of *Emojinating* (see Chapter 11). It is important to mention that some aspects are easier to apply in a visual blending with layered images. For raster images, other aspects may be more appropriate.

15.3.3.4 *User Perception*

Despite the importance of all the topics already mentioned, the quality of the visual blend will always depend on user perception and interpretation, and a blend may be considered bad even if it is conceptually grounded. Martins et al. (2015) take into consideration criteria that can be used to define creativity – i.e. *novelty*, *surprise* and *value* (Boden, 2004) – and asks participants to evaluate visual blends based on the following topics: *Overall impression*; *Novelty/Surprise*; *Interestingness*; *Aesthetic appeal*; *Comicality/Humor*; *Coherence/Consistency*; *Evoques positive feelings*; *Evoques negative feelings*; and *Creative industries potential*.

These topics can be subjective and people may evaluate blends differently. Due to this, providing a way for the user to interact with the system allows changes to be made to the blends, making them more suitable to the users' preferences. A method used by some systems is Interactive Evolutionary Computation (IEC) (as is the case with *Emojinating*), which consists in including the user in the task of fitness assignment and evolving solutions that match their preference.

15.3.4 *Elaboration*

A big part of the conceptual process may occur after the visual blending is done – consisting of an elaboration. This elaboration and consequent interpretation may in turn serve to provide justification for the previously done visual blend and also as a way to improve it – resulting in a return to a previous stage for a new iteration.

15.3.4.1 *Naming*

One example of elaboration is the production of names. Pollak et al. (2015) present a prototype for name generation based on an investigation focused on the principles of creating lexical blends based on visual blends (blended animals). Pollak et al. (2015) identify the following mechanisms used in name formation: L1-concatenation blends; L2-portmanteaux (e.g. "rabbear" for rabbit and bear); L3-blending based on visible characteristics; L4-blending using background knowledge and L5-bisociative blends (e.g. "mickey" the bear for mouse and bear). These mechanisms may also be used for blends that do not use animals.

15.3.4.2 *Descriptions*

In addition to names, there is also the potential to produce descriptions based on the visual blend. Techniques such as image captioning (Feng et al., 2019) may be used for this purpose. Ideally, a system that produces descriptions could produce an elaboration on the context of the blend. For example, mixing two animals leads to questioning the context of the hybrid animal: *Where does it live? What does it eat? How does it behave in relation to other animals?* All these questions would need to be addressed using a process of conceptual blending by getting characteristics from the two mental spaces. An example can be observed in the concept *clown fish*: does it live in the sea and looks like a clown or does it live in a circus and looks like a fish? Obviously, one of the situations has a higher likelihood, which makes it more plausible; but the surprising nature of the other option makes it so that in terms of creativity it has more potential.

Moreover, a creative system would have great advantages in providing the user with explanations for the produced blends. The descriptions can be seen as such and used to make the process of blending clearer to the user (Cook et al., 2019).

15.3.5 *Other aspects*

Having presented the four stages, we now address a set of aspects that, in our opinion, will be key in implementing a general model for visual conceptual blending.

15.3.5.1 *Modularity*

Most of the systems described before work in an individual way with no connection to others. An exception is *Vismantic* (Xiao and Linkola, 2015), which is integrated into a platform for workflow management – *ConCreTeFlows*. Martins et al. (2019a) focus on this platform and present an example of how it can be used to develop Computational Creativity (CC) software components that can be shared, used and reused to produce complex computational pipelines. We believe that an implementation of a general model for visual conceptual blend will benefit from using such a modular approach, allowing multiple users to contribute to the system.

15.3.5.2 *Multi-approach*

In addition to having several modules that deal with different tasks, as we have seen earlier, there are several methods that can be employed for each of the tasks (e.g. conceptualisation can be based on perceptual features, affordances, etc.). The suitability of these methods often depends on the type of problem at hand (i.e. the characteristics of the blend) and, as such, no optimal approach exists. A solution to this multi-approach situation is to follow a similar strategy to the one presented by Cardoso et al. (2015) – using a global workspace and a number of components that compete for access to it. Each component could be seen as an agent. At each time, the agent that is able to produce the most relevant output is given access to the workspace. This would consist in having solutions being produced by each of the agents and finding the best.

15.3.5.3 *User-centred*

The quality of a visual blending always depends on user perception, thus being of open-ended nature. As such, the user should be viewed as having a central role. The modular approach suggested earlier is dependent on having a user interact with the platform to build a pipeline of components. We go one step further and propose that the user should also have an active role in producing the visual blends.

First, the interaction with the user has great potential to be explored as it can be used to iteratively improve the quality of the blend, both visually and conceptually. This would have an effect on which approach is used at each task, depending on the user evaluation. Moreover, the user would guide the blend production in terms of improving second-order features (e.g. colour) or even extending the conceptual reach when no blends could be produced with the existing knowledge.

Another possibility is to provide the user with a way of selecting the creativity degree – e.g. low creativity resulting in more literal representations and high creativity in more non-literal results.

15.4 SUMMARY

In this chapter, our goal was to take a step closer to outlining a model for visual conceptual blending that can be instantiated in a fully operational computational system. We argued that a visual conceptual blending process should not only result in a visual blend produced for a given concept but also be complemented by a conceptual layer developed through elaboration. We proposed a roadmap for the production of visual conceptual blends. This roadmap can be instantiated into a modular system, in which the different stages of blend production occur in an iterative manner, allowing the user to go back to improve the blend and its elaboration. Nonetheless, the roadmap is only a step towards a formalisation of visual conceptual blending and should not be seen as a closed proposal, as it will benefit from future iterations.

In the following chapters, we focus on specific questions related to the proposed roadmap, such as the use of affordances in the production of visual blends (Chapter 16) and the development of an emoji categorisation oriented towards visual blending (Chapter 17).

16

IMAGE SCHEMAS

Computational systems that produce visual representations of concepts mostly focus on perceptual characteristics and overlook conceptual ones (e.g. *affordances*). The work that we have presented in previous chapters is also no exception to this. Despite this, in the previous chapter, we have mentioned that affordances can be used to find blending possibilities and even guide the process of blending.

In this chapter, we analyse how affordance-related features can be considered in computational systems for the visual representation of concepts, through the use of *image schemas*. First, we deconstruct well-known used icons to show the role of *image schemas*, then we use examples to illustrate how visual representations can be produced using image schemas and discuss existing issues.

This chapter is based on the work described in the paper by Cunha, Martins, and Machado (2018c).

16.1 CONTEXT

Two types of categorisation processes can be said to take place in concept formation: *perceptual* and *conceptual* (Mandler, 2000). Perceptual categorisation has to do with perceptual features, i.e. what objects look like, whereas conceptual categorisation is related to purpose and usage, i.e. affordances (Hedblom and Kutz, 2015). When defining a concept (e.g. *house*), the perceptual features (i.e. how a house looks like) are not enough and conceptual aspects should also be considered (i.e. what it can be used for), as pointed out by Hedblom and Kutz (2015). Despite this, little importance has been given to conceptual processes in the domain of visual representation of concepts, with systems mostly focusing on perceptual features (shapes, colours, etc.).

Kuhn (2007) explored the idea that affordances can be modelled using image schemas. This notion has been used in the computational modelling of concept invention and conceptual blending (e.g. Hedblom, Kutz, and Neuhaus, 2016). Image schemas are learned spatio-temporal relations that can be seen as conceptual building blocks (e.g. CONTAINMENT). Although they are not visual by nature, several authors have used visualisations in order to make their ideas clearer to the reader. Some examples are: SOURCE_PATH_GOAL and EQUILIBRIUM (Johnson, 1987) (Fig. 16.1); eight different visualisations for CONTAINMENT (Bennett and Cialone, 2014); the PATH-FOLLOWING image schema family (Hedblom, Kutz, and Neuhaus, 2015); CONTACT, SUPPORT, VERTICAL-

Figure 16.1: Visual representation of SOURCE_PATH_GOAL (left) and EQUILIBRIUM (right), adapted from Johnson (1987).

Figure 16.2: Pictograms for *escalator*, *luggage trolley* and *ferryboat*

ITY and ATTRACTION (Hedblom et al., 2017); and MOVEMENT-ALONG-PATH (Besold, Hedblom, and Kutz, 2017).

These visualisations of image schemas are aligned with spatial relations used in visual blending, for example, *inside*(x, y), used in the *Pig, Angel and Cactus* experiment described in Section 6.2.1.2, or *above*(x, y) used by Confalonieri et al. (2015). However, in these examples, spatial relations are mostly used as an aid for element positioning. Confalonieri et al. (2015) blended computer icons, which were composed of signs (e.g. a magnifying glass) and spatial relations between them. Different meanings were attained depending on the combination of sign and relation (i.e. a *downwards-pointing arrow* could lead to both *download X* or *download-to X*, depending on the used relation). In the *Pig, Angel and Cactus* experiment, we focused on perceptual aspects and tried to produce visual blends by identifying the prototypical parts of concepts and using previously defined spatial relations.

We propose that, in addition to perceptual features (e.g. prototypical parts), affordances should be considered in systems for the visual representation of concepts. These can be modelled using image schemas, as suggested by Kuhn (2007). As such, the concept *house* can be represented using its prototypical parts (e.g. walls and roof) but also by focusing on its affordance of being used as shelter – i.e. to offer protection. The idea of using image schemas in the visual representation of concepts is also addressed by Falomir and Plaza (2019), who propose an approach to computationally model the understanding of conceptual blends by a receiver agent. Their approach is based on disintegration and decompression of input visual representations of novel concepts (e.g. blended icons) and consequent recreation of the blends, using qualitative spatial descriptors and image schemas. Despite the alignment with our work, marked by the proposal of image schema integration in processes related to the visual representation of concepts, the goal of Falomir and Plaza (2019) is different from ours. Whereas they address understanding (from form to content or meaning), we focus on generation (from meaning to form).

16.2 APPROACH

In addition to perceptual features, the affordances related to concepts can also be observed in pictograms of signage systems. For example, the potential use of an *escalator* is represented using an *arrow* (Fig. 16.2) and the idea of SUPPORT from a *luggage trolley* or a *ferryboat* is illustrated through the inclusion of the entity that they *support* – a suitcase and a car, respectively (Fig. 16.2). Moreover, other communication systems also make use of image schemas for the representation of concepts – e.g. see the use of SOURCE_PATH_GOAL in the distinction between *entrance/exit* or *start/arrival*, shown in Fig. 2.9.

Having these examples as inspiration, we present an approach for the integration of affordances (using image schemas) in systems for the visual representation of concepts through visual blending. We believe that image schemas can be used to guide the process and validate the results, minimising the number of "nonsense" solutions, as argued by Hedblom, Kutz, and Neuhaus (2018).

In this section, we first explain the approach and then we give some illustrative examples to show the potential of considering conceptual aspects in visual (conceptual) blending.

16.2.1 *A 4-step pipeline*

The proposed approach uses the following 4-step pipeline:

1. **Identification of the concept:** The first step consists in identifying the concept to be visually represented. Computational approaches to the visual representation of concepts often allow the user to freely introduce concepts. For example, *Emojinating* (Chapter 8) takes as input single-word (e.g. *bank*) and double-word concepts (e.g. *mother ship*).

2. **Identification of image schemas:** This step consists in identifying image schemas related to the concepts, which is a challenging task. We identify several methodologies that can be used for the identification of image schemas. The methodology presented by Kuhn (2007) uses *WordNet* glosses to extract image schematic structures for concepts (e.g. identifying CONTAINMENT for *house*). The gathering and analysis of example sentences for each concept would allow the identification of possible image schemas related to them – matching *human habitation* and *living quarters* with the idea of "containing humans" (see the *house* descriptions in Fig. 16.3, retrieved from the *Oxford Dictionaries*[1] and *WordNet*[2]). Other approaches focus on the extraction of spatial descriptions

1 en.oxforddictionaries.com
2 wordnet.princeton.edu

from text. One example is the Generalised Upper Model ontology (GUM) by Bateman et al. (2010), which facilitates mappings between natural language spatial expressions and spatial calculi – using the preposition "on" indicates SUPPORT (e.g. "the suitcase is on the luggage trolley" or "the car is on the ferryboat") and using "in" indicates CONTAINMENT (e.g. "the suitcase is in the car" or "the car is in the garage"). In this model, SUPPORT and CONTAINMENT are seen as subconcepts of "Control", which is itself a subconcept of "FunctionalSpatialModality".

3. **Gathering input visual representations:** In the case of *house*, two input visual representations would be needed for its visual representation – the pictograms for *building* and *person* (as shown in step 2 of *house* in Fig. 16.3). Following the same strategy as the one used with *Emojinating*, a dataset of visual representations and corresponding semantic information could be used. Such a dataset allows matching concepts to visual representations (e.g. the word *baby* used as input to *Emojinating* system leads to the automatic retrieval of the baby icon shown in Fig. 16.4.

4. **Production of visual representations:** The last step concerns the use of the gathered visual representations (e.g. *building* and *person*) in combination with the identified image schema(s) (e.g. CONTAINMENT) to generate visual representations of the concept (e.g. *house*). This process of generation has several implementation issues of considerable complexity (positioning of elements, image schema activation, etc.), which we will describe in more detail in Section 16.3.

16.2.2 *Illustrative examples*

In order to show the potential of considering image schemas in systems for visual representation of concepts, we start by presenting three examples of icons from signage systems that show how image schemas are used in icon design (Fig. 16.3).

The first example is the icon for the concept *house*. The concept *house* can be represented using only perceptual features (e.g. the icon shown in step 1 only represents the roof and the walls of a *house*, see Fig. 16.3). However, it can also use the affordance of serving as a shelter. In this sense, it is important to mention that the roof shape may also be seen as affordance-indicating and not purely perceptual. By considering the affordance of serving as a shelter, one may relate it to the CONTAINMENT image schema (Kuhn, 2007) – identified in the example descriptions ("human habitation" or "living quarters"). The CONTAINMENT schema implies a *container entity* and a *contained entity*, which can be respectively linked to "building" or "dwelling", and "human", based on the

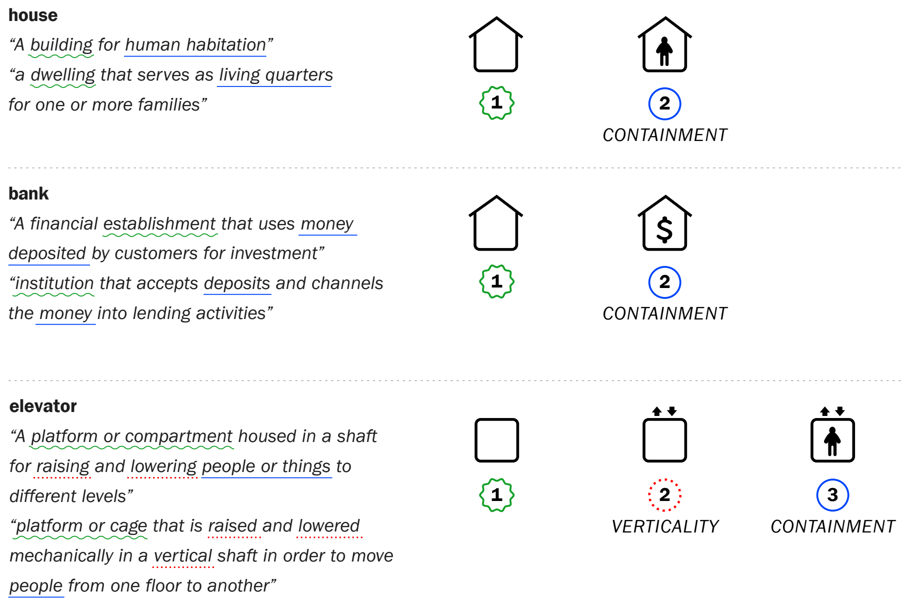


Figure 16.3: Identification of image schemas for *house*, *bank* and *elevator*, using examples retrieved from *Oxford Dictionaries* and *WordNet*. The different types of underline identify the steps taken in each example (best viewed in colour).

descriptions provided. This can result in a *person* sign placed inside of a *building* (see step 2 of *house* example in Fig. 16.3).

If we consider the concept *bank*, we reach a similar situation to *house*. Based on the action of "depositing" from the descriptions, we can also establish a connection to the CONTAINMENT image schema. This connection is further reinforced if we take into consideration other examples, such as the sentence "a bank account may contain funds, and if it is empty we can put some additional funds into the account and take them out again later" presented by Hedblom, Kutz, and Neuhaus (2018). As we already mentioned, the CONTAINMENT schema associates a container with something that it contains – in the case of bank and based on the descriptions, these two entities can be matched with "establishment" or "institution", and "money", respectively. As such, a possible representation can be a *building* sign that has a *dollar* sign inside (see the *bank* icon in Fig. 16.3).

A third example is the concept *elevator*, which is more complex as it deals with a combination of two different image schemas. The representation of complex abstract concepts using a combination of several image schemas is also addressed by Kuhn (2007). The main idea behind an elevator is its capability of moving upwards and downwards – based on the descriptions "for raising and lowering" or "raised and lowered mechanically" from Fig. 16.3. This can be translated into the VERTICALITY image schema, which is associated with movement. When dealing with static images, it can be represented using signs such as arrows (see *elevator* step 2 in Fig. 16.3). However, VERTICALITY is not the only image schema that can be associated with *elevator* – consider the question "what exactly does an elevator raise / lower?". Similarly to what happens with *house* and *bank*, *elevator* is also related to CONTAINMENT. From the descriptions in Fig. 16.3, one can identify that the contained entity for *elevator* is related to "people or things", which justifies the construction of the icon often used to represent *elevator*.

16.3 DISCUSSION

The examples analysed in Section 16.2.2 serve to show that there is potential in considering image schemas in the visual representation of concepts. Despite this, there are several issues regarding the implementation of the proposed approach. Moreover, the examples already presented (*house*, *bank* and *elevator*) are based on existing icons and, as such, they were analysed using a deconstruction method, which was performed at a very superficial level and avoided most of the existing issues. Using image schemas to generate novel visual representations is more complex than portrayed in the given examples.

In this section, we identify issues that have to be considered when using image schemas in a system for visual representation of concepts. The majority of the concepts used in the examples were collected from existing research work. We conduct a high-level analysis and interpretation of visual representations. Nonetheless, it is important to highlight again that decomposing visual representations into meaningful elements in visual perception is a complex process. For further reading on the topic, we refer the reader to Bateman, Wildfeuer, and Hiippala (2017), Black et al. (2017), Engelhardt (2002), and Tufte (1997).

16.3.1 *Image Schemas: Identification*

Regarding image schema identification, one of the issues is that not all concepts can be associated with image schemas and, as such, this approach will not work in every situation. In fact, for the visual representation of some concepts, perceptual features are more important than conceptual ones (e.g. *dog*). Moreover, the actual identification of an image schema from text is complex and a subject of study itself – e.g. words related to CONTAINMENT (Bennett and Cialone, 2014) and extraction of spatial descriptions from text (Bateman et al., 2010). Several approaches can be explored to identify image schemas, e.g. the use of metaphors associated with the concept being represented (we use this approach in examples given in the following sections).

16.3.2 *Image Schemas: Visual Representation*

Putting aside the identification of image schemas and focusing on their usage, there are some questions that need to be addressed. First, using

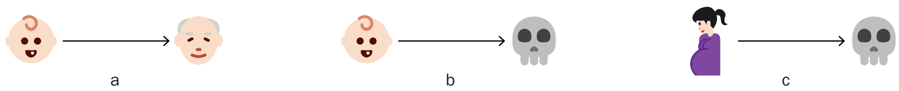


Figure 16.4: Three visual representations for *life* based on the metaphor *life is a journey* (Hedblom, Kutz, and Neuhaus, 2015), using different entities for SOURCE (*baby* or *baby in womb*) and GOAL (*old person* or *death*).

image schemas in the visual representation of concepts assumes that image schemas have a visual representation themselves. Despite this being true for some – easier to represent (e.g. SOURCE_PATH_GOAL) and even in line with spatial relations used in visual blending systems, e.g. CONTAINMENT, (Confalonieri et al., 2015) – others are much more complex and may not be so straightforward in terms of visual representation (e.g. EQUILIBRIUM in Fig. 16.1). As such, further study is required to identify the image schemas more suitable for visual representation.

In addition, schemas that can be considered simple may end up having an application more complex than initially expected. For example, CONTAINMENT only requires two entities which are combined using an inclusion relationship. Despite this, issues may arise when combining these entities – this example will be further detailed in a later section using the concept *mother ship*.

16.3.3 *Image Schemas: Entities*

Other image schemas regarded as simple may require extra signs in addition to the entities in order to be fully represented. The image schema SOURCE_PATH_GOAL, for example, can be visually represented using two entities (A and B) connected by an arrow, which indicates a transition between two points (Fig. 16.1). To use this image schema in the visual representation of concepts, two entities need to be identified – A, the *source*, and B, the *goal*. This identification is not always easy and may lead to different meanings, depending on the entities chosen. Consider, for example, the three representations for the concept *life* based on the metaphor "life is a journey" (Hedblom, Kutz, and Neuhaus, 2015), as shown in Fig. 16.4. First, the metaphor associates *life* to the image schema SOURCE_PATH_GOAL. As such, the two entities need to be identified and several possibilities exist. The first one (solution *a* in Fig. 16.4) consists in considering the SOURCE as the initial stage of life (infancy represented by a *baby*) and the GOAL as the last (old age represented by an *old person*). Despite being a possible solution, if we consider the GOAL as the end of life, it is more correct to choose an entity that represents death (portrayed using a *skull* in solution *b*). Similarly and in order to be exact, the beginning of life is when the baby is still inside the mother's womb, which can be represented by assigning a *pregnant woman* icon to the SOURCE (solution *c*). This example shows that for the same concept, based on the same metaphor, and using the

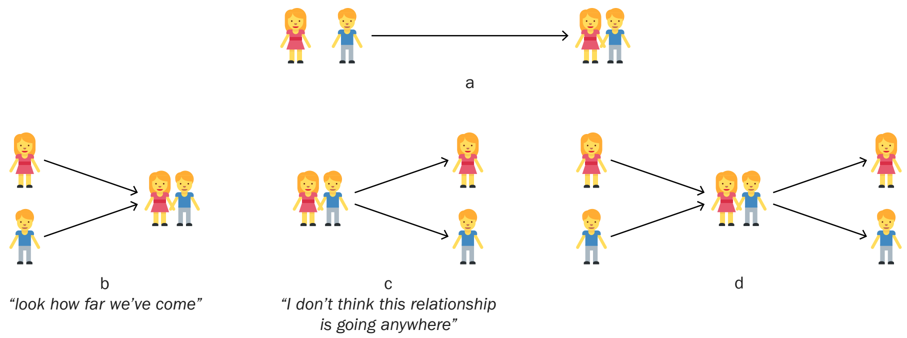


Figure 16.5: Four visual representations for *love* based on the metaphor *love is a journey* (Hurtienne, 2009), using different examples.

same image schema, several possibilities exist in terms of representation. This variety may also lead to different meanings – solutions *a* and *b* represent the development of the baby, whereas solution *c* can instead be interpreted as the progression of the mother towards death.

16.3.4 *Image Schemas: Concepts and Descriptions*

On the other hand, the application necessities of one image schema may change depending on the concept being represented. For example, changing the concept from *life* to *love* but maintaining the metaphor "is a journey" (Hurtienne, 2009) leads to the same SOURCE_PATH_GOAL image schema. The representation *a* of Fig. 16.5 follows the same procedure as the one used in *life* and consists in the SOURCE being two people separate and the GOAL two people holding hands. However, if we consider that the emphasis of journey is the path of each individual towards a state in which they are together, it might make more sense to represent the individual paths – *b* in Fig. 16.5 – which is different in terms of representation. Moreover, the representations *a* and *b* are based on the assumption that the journey is the path towards being together – which might be based on the description "look how far we've come" – but using a different description (e.g. "I don't think this relationship is going anywhere") may lead to the exact opposite – as seen in *c* of Fig. 16.5. There is even the possibility to use the two descriptions together, which represents the "journey" from two people from being separate to being together and ending up going separate ways again (*d* in Fig. 16.5). In this last example, a middle point is added to "the journey", increasing the complexity of the image schema application. These examples serve to show that the application requirements may vary, even using the same image schema and the same concept.

The use of different descriptions for the same concept may also lead to different image schemas, which completely changes the visual representation. For example, *love* can also be represented by using the metaphor "as unity" (Hurtienne, 2009). This metaphor infers that there

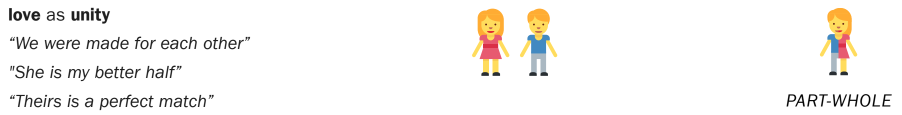


Figure 16.6: Visually representing *love* using the metaphor *love as unity* (Hurtienne, 2009) – examples used (left), initial visual representations (middle) and visual representation for *love* (right).

are two parts that make a whole, which leads to the part-whole image schema (see examples in Fig. 16.6). This image schema has a visual representation entirely different from the source_path_goal – two entities are now seen as parts from a whole. In source_path_goal the visual representation was more or less intuitive, whereas in part-whole it is not so obvious. One possible way to represent part-whole consists of the following procedure: (i) identify the entities and gather their visual representations (middle of Fig. 16.6); (ii) conduct a visual transformation to make them be seen as "parts" (e.g. cutting them in half); and then the parts can be put together to make one single entity (right side of Fig. 16.6). However, the transformation used may not work in every situation and the final result might not have an easy interpretation.

16.3.5 *Blending: Image Schema Activation*

The blending process aims to represent the meaning of the concept, which requires (i) the correct usage and the activation of the image schema(s), achieved by (ii) a correct combination of the input visual representations. In the previous example (*love as unity*), we already addressed issues that concern how image schemas can be activated in visual blending – transforming the input visual representations (e.g. cut in half) and afterwards merging them into a single element in order to activate the part-whole image schema.

Even using a simple image schema, e.g. containment, its activation may prove to be problematic. The containment image schema can be represented by one of the entities being placed inside the other. Consider for example the visual representation of two concepts – (1) "being inside of a boat" and (2) "being inside of a car" – using input visual representations (a *person*, a *boat* and several versions of *car*, see Fig. 16.7). One initial attempt to represent the two concepts might be to use the bounding box of the container entity's visual representation for placement of the contained entity (row a, Fig. 16.7) . However, this approach is not guaranteed to work and may lead to unwanted and even opposite meanings – "swimming / drowning" (*boat*), "being outside of / next to a car" (*car 1* and *car 3* activate the in-out image schema) and "being run-over" (*car 2*).

Another approach may be to only consider part of the visual representation (e.g. only considering the boat and excluding the water),

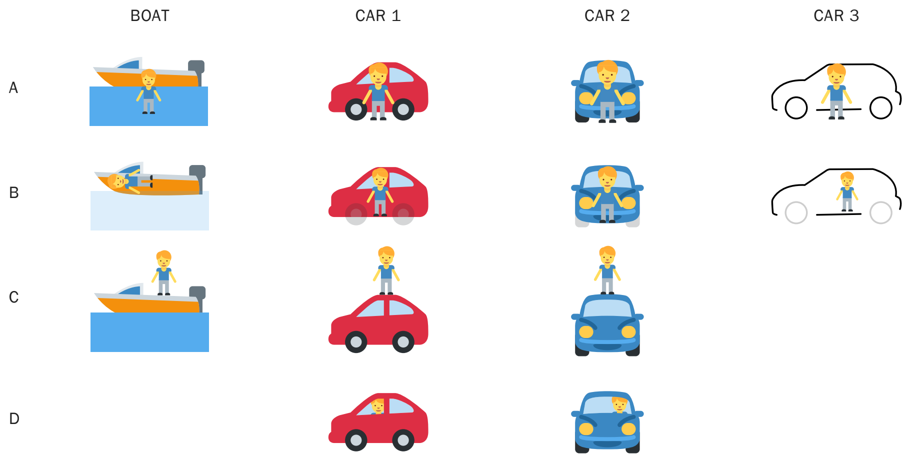


Figure 16.7: Experiments with CONTAINMENT image schema using different versions of *boat* and *car*.

use its bounding box for placement and apply the necessary transformations (e.g, rotation or scale) in order to place the contained entity inside it (see row B, Fig. 16.7). In addition to being dependent on context knowledge (knowing which parts to use), this approach only works in some cases (*car 3*) and may lead to incorrect solutions in others – *car 1* and *car 2*.

A solution for "being inside of a boat" can be achieved by placing the *person* on top of the *boat* (boat C). Although this works for *boat*, using it with the *car* will not activate the CONTAINMENT image schema (*car 1* and *car 2*) and may even activate other image schemas (e.g. UP-DOWN). In the case of the concept "being inside of a car" using the input visual representations *car 1* and *car 2*, the CONTAINMENT image schema is only activated by considering visibility aspects, thoroughly adjusting the layer order and placing the *person* behind the *car* structure (see row D). Such adjustments are, however, complex to implement in an automatic computational system, as they require context knowledge of the concept and depend on the input visual representations.

The subject of image schema activation is studied by other authors. Hurtienne (2009), for example, highlights the importance of "functional geometry" (combining the "appropriate" objects in the spatial scene) in image schema activation.

16.3.6 *Blending: Combination*

Having addressed the issue of activating an image schema, we now focus on the combination of entities to achieve a given meaning. Consider, for example, the concept *mother ship*, addressed by Hedblom, Kutz, and Neuhaus (2016), which is highly related to the CONTAINMENT image schema. The combination process is far from simple, even assuming that, for the visual representation of a given concept (e.g. *mother*

Figure 16.8: Input visual representations for *mother*, *baby* and *ship* (left) and three visual representations for *mother ship* (right).

ship), the adequate image schema is identified (e.g. CONTAINMENT), suitable entities are chosen (e.g. *mother*, *baby* and *ship*, see Fig. 16.8) and the system has knowledge of how to correctly apply the image schema (e.g. placing the *contained entity* inside the *container entity*). The first issue concerns the assignment of the *container* and the *contained* entities. For *mother ship* this is not trivial as both *mother* and *ship* can be seen as containers (*mother* may "contain" a baby and *ship* may "contain" cargo). As such, different interpretations will lead to different solutions (*a* and *b* in Fig. 16.8), which may not be considered valid – *a* can be considered as nonsense and *b* may lead to other meanings (e.g. the ship that carried Superman to earth when he was still a baby).

The solutions *a* and *b* were produced in a process of visual blending that only considered conceptual aspects from the individual entities (*mother* and *ship*), using the spacial relation inside to represent CONTAINMENT. In these solutions, the combination was performed without regarding conceptual aspects from the concept (*mother ship*), resulting in two "nonsense" blends (a ship baby inside a human mother and a human baby inside a ship mother). Although this may lead to possible solutions in certain situations, *mother ship* can be seen as a conceptual blend between the input spaces *mother* and *ship* and its visual representation should take this aspect into consideration. The idea behind the concept *mother* is not only of CONTAINMENT but CONTAINMENT of individuals of the same class – the mapping between *mother* and *ship* as *mother ship* is a ship that contains other ships (*c* in Fig. 16.8).

16.4 SUMMARY

Computational systems that generate visual representations of concepts often focus on perceptual features. In this chapter, we highlighted the potential of considering affordances in such systems. We described a possible approach for their integration in systems such *Emojinating* (see Chapter 8), based on the detection of image schemas related to the concepts being represented. We presented several examples to show the importance of image schemas in the design of visual representations (e.g. *icons*), identified issues that need to be addressed when implementing the proposed approach and compared solutions in terms of validity.

In this chapter, we made the following contributions: (i) the outline of an approach to include affordances in a system for visual (conceptual) blending, (ii) the analysis of a set of illustrative examples and (iii) the identification of implementation issues that should be addressed in the future.

The main implementation issues concern: (i) the identification of the image schemas, (ii) the visual representation of image schemas, (iii) the choice of the adequate entities, (iv) the meaning variation triggered by using different examples, (v) image schema activation and (vi) suitable combination of elements in the blending process. Despite all these issues (and others that may also exist), we believe that there is potential in the proposed approach for improving existing systems that visually represent concepts.

17 A BLENDING-ORIENTED CATEGORISATION

In Chapter 15, we outlined a roadmap for the development of visual conceptual blends. Among other aspects, we highlighted the relevance of the type of input objects in the process of blending. An example given was the one of the blends *dinosaur+park* and *dinosaur+fish*, in which the latter blend involved a mapping between similar parts (e.g. *heads*) whereas the former did not (Fig. 15.6). In addition, the role of similarity in the process of visual blending was also pointed out in previous chapters – in Section 4.2.3 we describe how *perceptual similarity* can be used to trigger a comparison mindset and in Section 5.3.2.4 we describe how *alignment* (*perceptual* and *conceptual*) is used in more than half of the analysed visual blends from the *Emoji Kitchen* dataset. In general, these two aspects (*conceptual categorisation* and *alignment*) should be taken into account when producing visual conceptual blends.

Throughout the thesis, we have argued that emoji are especially suitable to be used as input in visual blending (see Section 7.6.3 for more details). We explored this position by using them in most of the described case studies (e.g. *Emojinating* in Chapter 8 and *Moody Flags* in Chapter 14). In this chapter, we describe the development of an emoji categorisation oriented towards visual blending. With this categorisation, we aim at producing another visual blending resource. A first version of the categorisation was presented in Chapter 10 as it was used to obtain a measure of *visual concreteness* in the study therein described (see Section 10.2.3) and published in the publication by Cunha et al. (2020b). In this chapter, we describe the improvements that were made to the categorisation and validate it with a user study.

In Chapter 5, we conducted an analysis of two datasets that included visual blends. One of the datasets was *Emoji Kitchen*, which is composed of blends between emoji. In our analysis, we identified common transformations used to produce the blends. In the last part of this chapter, we conduct a second analysis of the images of *Emoji Kitchen* using the developed categorisation with the goal of understanding how categories can help to identify possible transformational patterns, which may be useful for visual conceptual blending purposes.

17.1 CONTEXT

In a process of visual conceptual blending, it is useful to know the type of objects being used in the blending, as well as assess similarities to other objects. When it comes to using a dataset of source images, as is the case with emoji, this sort of information can be gathered by previously conducting an analysis of the images and categorising them.

Concerning emoji, several categories have been in use. First, *Unicode* has a block system,[1] in which each block corresponds to a contiguous range of numeric character codes and has a unique name. A given block may also be subdivided into more specific groups. Each emoji has a code point that belongs to a specific block. For example, the emoji for *dove* 🕊 (`U+1F54A`) belongs to the "Religious symbols" subgroup of the "Miscellaneous Symbols And Pictographs" block. By analysing the blocks and subgroups, we conclude that the block level is too general (e.g. "Miscellaneous Symbols" and "Miscellaneous Technical") and the subgroups are too specific (e.g. "Ballot symbols").

In addition to this block system, *Unicode* also assigns each emoji to a category. At the time of writing, the emoji set (v14.0)[2] is organised into ten categories, which are subdivided into a total of 100 subgroups. We consider that the subgroups are too specific – e.g. animal-related emoji are divided into *amphibian*, *bird*, *bug*, *mammal*, *marine* and *reptile*.

Several authors use emoji categorisations, which employ multiple organisation criteria and are oriented towards different purposes, for example: Cappallo, Mensink, and Snoek (2015) use categories originated from a context of object recognition, more specifically the label categories of the MSCOCO dataset (Lin et al., 2014);[3] *Emojipedia* uses a thematic approach (Burge, 2013), which is also the categorisation used in the keyboards of *Apple*'s iOS and *Google*'s *Android*; Barbieri, Ronzano, and Saggion (2016) identify categories using a clustering approach; Vidal, Ares, and Jaeger (2016) use six categories based on the ones from *Emojipedia*; and Donato and Paggio (2017) create a categorisation that separates *events* and *activities* from *entities*, composed of a reduced number of broad categories with high recognisability. Bai et al. (2019) present a categorisation divided according to three emoji functions: *content* (uses the same categories as *Emojipedia*), *meaning* (behavioural vs non-behavioural) and *emotion* (positive, neutral and negative).

The majority of these categorisations tend to be either too general – most do not separate, for example, *vehicles* from *places* or *food* from *drinks* – or too specific – for example, Cappallo, Mensink, and Snoek (2015) over-specifies with a category related to *kitchenware*. Moreover, they mostly focus on *thematic similarity*, which does not fully cover our needs. Based on this analysis, we came to the conclusion that existing categorisations are not oriented towards visual blending and the development of such a categorisation would be useful for the implementation of visual blending systems.

---

1 `www.unicode.org/charts/nameslist/`

2 `unicode.org/emoji/charts/emoji-ordering.html`

3 Cappallo, Mensink, and Snoek (2015) states that 37 out of the 80 label categories from the MSCOCO 2014 release (Lin et al., 2014) are represented in the set of emoji. For our purposes, we resort to the list of supercategories of MSCOCO.

Table 17.1: Identification of common topics among different emoji categorisations. The table shows the categorisations proposed by *Unicode*, *Emojipedia*, Cappallo, Mensink, and Snoek (2015), Vidal, Ares, and Jaeger (2016), Barbieri, Ronzano, and Saggion (2016) and Donato and Paggio (2017). On the bottom are listed the categories that do not directly match a common topic. Note that the categories listed on the bottom and marked with * can be seen as belonging to the topic "Objects".

	Unicode	**Emojipedia**	**Cappallo**	**Vidal**	**Barbieri**	**Donato**
Food	food & drink	food & drink	food	food & drinks	eating & drinking	eating & drinking
Drinks	food & drink	food & drink		food & drinks	eating & drinking	eating & drinking
Animals	animals & nature	animals & nature	animal		sports & animals	nature & animals
Nature	animals & nature	animals & nature		nature	nature	nature & animals
Vehicles	travel & places	travel & places	vehicle	travel & places		traveling & commuting
Activity	activities	activity		activity		other activities
People	people & body	smileys & people	person			people
Places	travel & places	travel & places		travel & places		places
Objects	objects	objects	*	non-food objects		
Sports			sports		sports & animals	sport
Symbols	symbols	symbols			barber & symbols	
Music					music	music
Flags	flags	flags				
	component		*accessory	celebration	letters	events
	smileys		*appliance		free time	feelings
	& emotion		*electronic		unclear	
			*furniture		love & parties	
			*indoor		sad & tears	
			*kitchen		body gestures	
			*outdoor		& positive	

17.2 DEVELOPMENT OF A CATEGORISATION

For developing a categorisation to be used in visual blending, we identify two aspects to consider: (i) *thematic* (i.e. *beer* and *wine* are both *drinks*), and (ii) *visual similarity* (e.g. *faces* will likely have a similar structure).

Regarding thematic, we analysed existing categories and we identified the common topics by comparing category labels (see Table 17.1). The most frequent topic is "Food", which is present in all analysed categorisations, followed by "Drinks", "Animals", "Nature" and "Vehicles", which are present in all but one. An interesting perspective is the one by Donato and Paggio (2017), who designed their categorisation in a way that it keeps *events*, *activities* and *entities* separate. Such an approach is also suitable for visual blending purposes.

On the other hand, it is also important to have a categorisation that takes visual similarity into account. The emoji set can be seen as a visual language system in which a particular style is maintained across all characters, giving a sense of coherence. It is common that conceptually similar representations are also visually similar – e.g. the representations of people's and animals' faces have a similar structure. Moreover, in these cases, elements that are conceptually similar are usually located in similar positions (e.g. *eyes* and *mouth*). For this reason, categories should also aim to group emoji that are visually similar. This sort of grouping based on rendering similarity is also done by *Unicode*, as pointed out by Donato and Paggio (2017). Examples of this are the subgroups *hand-fingers-open* (hands with open fingers) and *hand-single-finger* (hands with a pointing finger).

Considering that *Emojinating* uses data from *EmojiNet* (Wijeratne et al., 2017b), our initial approach was to take advantage of the categories from the *EmojiNet* dataset, which uses a combination of the *Unicode* block and subgroup as a category. Using this combination of block and subgroup, *EmojiNet* has a total of 117 different categories, including emoji of a total of 21 different blocks. However, upon analysing the *EmojiNet* dataset, we came to the conclusion that several emoji did not have any information regarding category (401 out of 2389) – only a "null" value could be found. Many of these "null" emoji are flags or resultant from modifiers, which should adopt the same category of the default version. Additionally, despite using criteria that are more or less in line with ours (*thematic* and *rendering similarity*), the categories are not well-fit to be used for blending purposes and, in some cases, are too specific, leading to a high degree of overlay (e.g. emoji which we consider *Animals* were in four different categories).

Due to these issues, we developed a categorisation, based on three criteria: (i) distinction between *entities*, *objects* and *places*; (ii) grouping according to thematic – e.g. *Animals* (entity type) or *Food* (object type); (iii) grouping according to visual characteristics and similarity

Table 17.2: Blend-oriented categorisation version 1. The table shows the category name, number of emoji (#E), value of visual concreteness assigned (C) and description. We considered the emoji included in *EmojiNet* (Wijeratne et al., 2017b), which is aligned with *Unicode Emoji 4.0*.

category	**#E**	**C**	**description**
Activities	12	4	daily activity
Animals	58	5	animals
Arrows	24	1	arrow-like shapes
Astronomy	23	5	celestial objects (e.g. moon)
BodyPart-arms	18	5	shows human limb from the shoulder to the hand
BodyPart-face	18	5	body parts belonging to the face
BodyPart-fingers	6	5	shows fingers but not the whole hand
BodyPart-hands	157	3	shows hand doing an action or gesture
Buildings	23	4	shows structure that has walls and may have a roof
Clothing	12	5	objects associated with clothing and accessories
Clothing-body	7	5	wearable on the top part of the body
Clothing-head	7	5	wearable on the head
Drinks	11	5	drinkable items
Faces	76	3	yellow faces showing emotions
Faces-animal	24	5	show animal faces in frontal view
Faces-cat	9	3	cat faces showing emotions
Faces-monkey	3	3	monkey faces doing gestures
Faces-other	14	5	shows face of entity, frontal view
Faces-people	97	5	shows face of person, frontal view
Flags	262	2	shows a flag in frontal view without pole
Food	72	5	edible items
Geometric Shapes	31	1	shows a geometric figure
Letters or punctuation	47	2	shows letters or punctuation marks
Lines	4	1	shows a long, narrow mark or band
Objects	248	4	shows a material thing that can be seen and touched
Objects-outside	29	4	objects found outside
People	59	4	people in several configurations
People-frontal	34	5	people, frontal view with shoulders
People-fullBody	42	4	full body of person
People-gesture	126	3	people doing gestures
People-role	331	4	people in different roles, frontal view with shoulders
Places	23	4	shows a location
Plants	23	5	living organisms lacking the power of locomotion
Sports	234	4	people practicing sports
Symbols	91	1	signs that represent something
Symbols-Pictorial	58	3	signs that represent something and are pictorial
Vehicles	51	5	conveyances that transport people or goods
Weather	25	5	weather related objects

– e.g. *Faces-animal* are kept separate from *Animals*. This third criterion is mostly focused on the distinction between different blend types: e.g. *Faces* may be more appropriate for *fusion* (exchange of parts) and *Clothing* (e.g. shoes) more suitable for *juxtaposition*. Ideally, emoji that share similar features and style should be grouped.

The initial category list was based on the common topics, identified in Table 17.1. The categorisation went through a process of iterative improvement, in which five independent evaluators were consulted until full agreement was reached. Based on their analysis, some categories were added (e.g. *Plants* and *Weather*). The final list of the categories can be seen in Table 17.2. As already mentioned, this first version of the categorisation was developed in the context of our experiments with *Emojinating*, described in Chapter 10.

One of the main applications for this categorisation is its use for visual blending purposes, helping with the identification of the best blend type to use based on the input emoji. In addition to this, we used the categorisation to obtain a metric of visual concreteness. Visual concreteness is defined as comprising two dimensions (Prada et al., 2016): (i) *concreteness* – "stimuli that (...) refer to objects, materials or people should be considered concrete (...) otherwise, they should be considered as more abstract" – and (ii) *meaningfulness* – "to what extent the stimulus conveys a meaning". This combination is in agreement with the results obtained by Prada et al. (2016), which positively correlated concreteness with meaningfulness – according to Prada et al. (2016) this result is congruent with the findings of other studies. As such, our visual concreteness value ranges from 1 (very abstract, symbolic or ambiguous in meaning) to 5 (very concrete and with obvious meaning). A similar study was conducted by Rodrigues et al. (2018), using a dimension of *Clarity* to classify 85 emoticons and 153 emoji – mostly faces, excluding the majority of emoji set. Each category was assigned with a value of visual concreteness, which was also iteratively improved based on the feedback from the five independent evaluators. For example, the categories *Objects* and *Objects-outside* had initially a visual concreteness value of 5 and were later reduced to a value of 4 as a great number of their emoji were pointed out as not being immediately perceivable, e.g. *ticket* 🎫 (`U+1F3AB`).

17.3 INITIAL IMPROVEMENTS TO THE CATEGORISATION

Although we consider the initial version of the categorisation to be a useful asset, we identify several aspects that could be improved. This first version has a total of 38 different categories (see Table 17.2), being based on the emoji list from *EmojiNet* dataset (Wijeratne et al., 2017a), which includes 2389 emoji (aligned with *Unicode Emoji 4.0*, released in 2016). To improve the categorisation, we produced a new version in which we considered *Unicode Emoji 13.0* (released March 2020) and used the *Twemoji* image dataset version 13.0.1. Despite the high number of categories in the initial version, we identified features in some emoji that required the creation of additional categories, some of which were due to newly added emoji. Two categories had to be divided to account for significant differences in the design of their emoji: the emoji that were initially in *Animals* were divided into *Animals-profile*, e.g. *giraffe* 🦒 (`U+1F992`), and *Animals-frontal*, e.g. *orangutan* 🦧 (`U+1F9A7`); and some of the emoji in *Faces-animal* were moved into *Faces-animal-profile*, e.g. *unicorn* 🦄 (`U+1F984`). In these two cases, the addition of new categories was motivated by the existence of *profile* and *frontal* perspectives in the emoji, which may have different needs in a visual blending process. Three other categories were created based on existing ones – body parts were already being categorised separately (e.g. arms in *BodyPart-arms* and hands in *BodyPart-hands*) so it made sense to create categories for *BodyPart-feet*, *BodyPart-hair* and *BodyPart-legs*. The list of categories of this second version can be see in Table 17.3. It is worth noting that the categorisation is based on emoji from *Twemoji* and that some image differences exist both across vendors (Miller et al., 2016), e.g. *penguin* (`U+1F427`) in iOS 🐧 and *Twemoji* 🐧, and also in different *Twemoji* versions, e.g. *giraffe* 🦒 in *Twemoji* 2.3.

17.4 VALIDATING THE CATEGORISATION

The main issues with the first version of the categorisation were that it was only evaluated by five people and the concreteness value was assigned to category and not to emoji. As there is variation among emoji belonging to the same category, the proper way of obtaining visual concreteness values would be to have several users classify each emoji in terms of concreteness. The category concreteness would then be calculated with the average of its emoji.

In order to evaluate the categorisation, we conducted a user study on the crowdsourcing platform *Appen* (former *Figure Eight*).

17.4.1 *Experiment Setup*

Our goal with this study is two-fold: (i) validate the category assignment; (ii) obtain a concreteness value for each emoji.

We conducted the testing of the categorisation in two stages. In the first stage, we used images from *Twemoji* 12.1.2 (referred to as 12). Upon the release of *Twemoji* 13.0.1, we decided to complement the testing with a second stage for it to be aligned with the latest version at the time. *Twemoji* 13.0.1 (referred to as 13) is composed of 3360 emoji images, adhering to *Unicode Emoji 13.0* spec.

Even though the number of emoji is currently more than three thousand, a great majority is the result of modifiers – e.g. emoji that change in skin colour or gender. These different versions, despite being worth studying, are not the focus of our work and testing only one version is sufficient. Moreover, 264 emoji are flags. We considered that testing every flag emoji was unnecessary as we expect that eventual differences in the results will be due to cultural knowledge, an aspect that we do not intend to address. We filtered out the variations and produced a subset of emoji to be included in the study, based on the following guidelines:

- use the "default" emoji (e.g.), which does not represent any skin tone (e.g.);
- use the version corresponding to "person" (gender-neutral). If the "person" version does not exist, we use the "man" one;
- only consider one flag per continent (Antarctica, Australia, Belgium, Brazil, Canada, Nepal, Mali) and also the European Union, rainbow and pirate flags;
- upon high similarity between emoji (the hours emoji group or the two magnifying glass emoji), only consider one (the *eight o'clock* emoji `U+1f557` and the *magnifying glass tilted left* `U+1f50d`, respectively).

In *Twemoji* v13, in addition to new emoji being added, some of the existing ones were updated and have now a different design, e.g. the *dagger* (`U+1F5E1`) vs , which means that the tests conducted in the first stage were outdated and could not be considered. Some gender-neutral emoji (e.g. *judge*) did not exist in v12 but have been added to v13. In these cases, the emoji tested in stage 1 is maintained in the tested set (now corresponding to *man judge*) and we tested the gender-neutral version (*judge*) in stage 2. Also, some of the images of version 12 for the *person* version were identical to the version of either *man* or *woman*, e.g. the image of *zombie* was equal to *man zombie* . In version 13, new designs are now used for some of these cases – e.g. a neutral version is now used for *zombie* (`U+1F9DF`). Given that some of these cases were tested in stage 1, instead of discarding the results, we use them for the emoji that uses the same image (e.g. results for v12 *zombie* are now used for *man zombie*). A total of 46 emoji images were reused for other emoji.

In stage 2, we conducted tests on the new emoji and also on the ones that have a different design. Moreover, from the preliminary results of stage 1, some of the emoji had inconclusive results (same number of participants selecting different answers). We decided to conduct more tests with these emoji in stage 2. A total of 1391 different emoji were tested and 23 emoji had a different version tested in stage 1, which could not be reused and therefore were discarded.

Overall, there is an average of 32.35 emoji per category with a standard deviation of 49.01. There are two categories with only one emoji tested (*BodyPart-feet* and *BodyPart-fingers*) since they only have one default emoji and five modifier ones (we only tested default ones). The category with more tested emoji was *Objects*, which had 300 tested emoji out of a total of 324.

Each emoji belongs to a single category. To design the experiment, for each category, we selected a closely related category, a slightly related category and a not related category. These were used as distractors. In addition, there was also the option of "none of the above", which was to be selected when the participant considered that the emoji did not belong to any of the categories. This setup concerned the first task: *selecting the category of a given emoji*. The possible answers were the right category, the three distractors and "none of the above". The second task concerned the *assignment of a visual concreteness value to the emoji*, from 1 (very abstract, symbolic or ambiguous in meaning) to 5 (very concrete and with obvious meaning).

To conduct the user study we used the *Appen* platform. We configured the experiment for English speaking participants.

17.4.2 *Results*

Each image (emoji) was seen by a minimum of 3 participants (average of 5.18, standard deviation 10.37 and mode 3), resulting in 7358 emoji evaluations. We had a total of 221 participants, with an average number of evaluations of 33.29 per participant, a standard deviation of 22.21 and a mode of 10.

In regards to the task of assigning a category to emoji, in 772 emoji all the participants selected the correct category and in 392 emoji it was selected by the majority. In 27 emoji, the correct category was selected an equal number of times as other answers. A wrong answer was selected by the majority of participants in 159 emoji, 35 of which by all the participants who tested the emoji. The categories with the highest number of cases of a wrong answer being selected by the majority were *Symbols* (25 cases) and *Symbols-Pictorial* (25 cases). The answer "none of the above" was selected by the majority of the participants in 40 emoji, 15 of which by all the participants. More than half of these cases happened with emoji of the *Objects* category (21 cases).

For each emoji, we calculated the percentage of each possible answer (correct, closely related, slightly related, not related and none). Then, for each category, we calculated the average for each of these percentages (Table 17.3). In terms of the average of the percentage of right answers, 13 categories have an average above or equal to 90%, 13 between 75-90%, 13 between 50-75% and the remaining four categories had between 25-50%. If we consider the sum of the averages of the right answer and the closely related one, we see results even better, for example, the group of above or equal to 90% increases to 26 categories and the 25-50% is reduced to one.

Table 17.3: Validation results of categorisation version 2. The table shows for each category the total number of emoji (#E), the number of emoji that are default (D), the number of emoji used in the study (T) and the average percentage of each type of answer – correct, closely related (“closely”), slightly related (“slightly”), not related (“not”) and none.

category	#E	D	T	correct	closely	slightly	not	none
Activities	12	2	2	0.83	0.00	0.17	0.00	0.00
Animals-frontal	24	24	24	0.70	0.24	0.03	0.00	0.03
Animals-profile	65	65	65	0.85	0.11	0.04	0.00	0.01
Arrows	24	24	24	0.95	0.05	0.00	0.00	0.00
Astronomy	24	24	24	0.99	0.00	0.01	0.00	0.00
BodyPart-arms	19	4	4	0.54	0.46	0.00	0.00	0.00
BodyPart-face	24	9	9	0.80	0.04	0.05	0.00	0.12
BodyPart-feet	6	1	1	1.00	0.00	0.00	0.00	0.00
BodyPart-fingers	6	1	1	1.00	0.00	0.00	0.00	0.00
BodyPart-hair	4	4	4	0.42	0.08	0.00	0.00	0.50
BodyPart-hands	181	31	30	0.95	0.02	0.01	0.01	0.00
BodyPart-legs	7	2	2	1.00	0.00	0.00	0.00	0.00
Buildings	26	26	25	0.88	0.08	0.03	0.00	0.01
Clothing-body	13	13	13	0.75	0.00	0.25	0.00	0.00
Clothing-head	10	10	10	0.70	0.03	0.27	0.00	0.00
Clothing	21	21	21	0.95	0.00	0.02	0.00	0.04
Drinks	15	15	15	0.69	0.17	0.14	0.00	0.00
Faces-animal-profile	6	6	6	0.88	0.08	0.04	0.00	0.00
Faces-animal	22	22	22	0.76	0.17	0.08	0.00	0.00
Faces-cat	9	9	9	0.97	0.00	0.03	0.00	0.00
Faces-monkey	3	3	3	1.00	0.00	0.00	0.00	0.00
Faces-other	14	14	14	0.62	0.05	0.01	0.01	0.30
Faces-people	210	33	26	0.60	0.33	0.01	0.06	0.01
Faces	94	94	94	0.93	0.04	0.03	0.00	0.01
Flags	264	264	10	0.86	0.00	0.11	0.00	0.03
Food	99	99	99	0.92	0.03	0.01	0.01	0.03
Geometric Shapes	43	38	20	0.84	0.00	0.15	0.00	0.00
Letters or punctuation	73	73	46	0.59	0.04	0.35	0.00	0.02
Lines	4	4	4	0.31	0.04	0.54	0.00	0.12
Objects-outside	34	34	34	0.56	0.21	0.10	0.06	0.06
Objects	324	324	300	0.83	0.06	0.03	0.01	0.08
People-frontal	35	6	4	0.94	0.00	0.06	0.00	0.00
People-fullBody	281	34	22	0.76	0.05	0.15	0.01	0.03
People-gesture	144	24	16	0.61	0.39	0.00	0.00	0.00
People-role	595	104	70	0.65	0.28	0.04	0.02	0.01
People	83	43	12	0.43	0.08	0.45	0.00	0.03
Places	23	23	23	0.67	0.20	0.04	0.00	0.09
Plants	25	25	25	0.99	0.00	0.00	0.00	0.01
Sports	256	46	25	0.87	0.11	0.02	0.00	0.00
Symbols-Pictorial	63	63	58	0.49	0.44	0.02	0.03	0.02
Symbols	93	93	93	0.67	0.30	0.02	0.01	0.01
Vehicles	57	57	57	0.89	0.03	0.04	0.01	0.03
Weather	25	25	25	0.76	0.08	0.01	0.03	0.12

In 40 out of all 43 categories, the right answer has the highest average percentage value. In two of the remaining three, the highest average percentage is obtained by the "slightly related" answer – 53% in *Lines* and 45% in *People*. In the case of *Lines*, in three out of the four tested emoji, the majority of the participants selected "Symbols: signs that represent something", which we assume was due to their interpretation of something more than a simple line. In the case of *People*, in six out of the 12 tested emoji, the majority of participants selected the slightly related option, "People-frontal: people, frontal view with shoulders". Upon analysing the emoji, we realise that the description of the category is also suitable and more specific than "People: people in several configurations" – all six emoji show people in frontal perspective with at least one visible shoulder. Despite this, we consider that these emoji are placed in the right category, as they either show more than one person (👪 `U+1F46A`), do not show both shoulders or involve an activity (e.g. *cutting hair* 💇 `U+1F487`). The distinction between these two categories serves a visual blending purpose by grouping similar postures. To fix this, the description of *People-frontal* was changed into "a person, frontal view showing both shoulders and head".

The third case in which the highest average percentage value was not obtained by the right answer happened with the category *BodyPart-hair*. In this case, the highest value was obtained by "none" (50%), which is the worst possible result. However, this result is easy to justify. The *BodyPart-hair* category only had four emoji tested, one of which is a yellow shape that represents a bald head (🦲 `U+1F9B2`). All participants selected "none" as an answer for the bald head emoji, which highly affected the results obtained.

Another category had a high average percentage of "none" – *Faces-other* obtained 30.1%. This result can also be explained. The *Faces-other* category has a total of 14 tested emoji and in 3 of them, the majority of the participants selected "none" (*sun with face* 🌞 `U+1F31E`; *goblin* 👺 `U+1F47A`; and *pile of poo* 💩 `U+1F4A9`). As an example, one of these emoji is the *pile of poo*, described by *Emojipedia* as "a swirl of brown poop, shaped like soft-serve ice cream with large, excited eyes and a big, friendly smile", thus fitting the category *Faces-other*. In spite of this, all participants selected "none", which negatively affected the results.

If we analyse the percentages of each possible answer of the categories with the highest number of cases of a wrong answer being selected by the majority (*Symbols* and *Symbols-Pictorial*), we notice that the right answer was only selected in 66.7% of the cases for *Symbols* and in 49.4% for *Symbols-Pictorial*. However, both categories had a high percentage of "closely related" being selected: for *Symbols* we used *Symbols-Pictorial* as closely related, which obtained 29.7%; for *Symbols-Pictorial* we used *Symbols* as closely related, which obtained 43.9%. The results show that there was a high uncertainty when it came to choosing between the two categories. In any case, for visual blending purposes, this result is not very relevant. Last but not least, the average of the percentage of "not related" reached its maximum of 6.23% in the category *Objects-outside*.

17.5 IMPROVING THE CATEGORISATION

In general, we consider that the results of the study described in the previous section show that the initial category assignment is appropriate in most cases. Despite this, we analysed the results of individual emoji to assess if changes were needed. For this, we focused on cases in which there was a high rate of wrong answers from the participants.

First, we noticed that the distinction between some categories was difficult to be perceived. Some examples are: *Clothing* vs *Clothing-body*, *Animals-frontal* vs *Animals-profile* and *Objects-outside* vs *Objects*. In most cases, we do not consider these difficulties as relevant because the purpose of these categories is related to visual blending (e.g. organising a frontal perspective in a different group from a profile one to facilitate the blending process). In some cases, the distinction is more related to thematic, e.g. *Objects-outside* category depends on the participant identifying a given object as normally being found outside.

There are also specific cases that are interesting to mention. For example, the emoji *pig nose* 🐽 (`U+1F43D`) is assigned to the category *Body-Part-face* but all participants selected "none" as an answer. Despite this, we consider that in a process of visual blending it may work as a face element (e.g. *eye* 👁 `U+1F441` or *mouth* 👄 `U+1F444`), thus we decided not to change it.

Nonetheless, some changes were necessary. For example, the *mushroom* 🍄 (`U+1F344`) was initially categorised as *Food* and, as all participants chose the category *Plants*, we decided to change its category. Similarly, the *kite* 🪁 (`U+1FA81`), which was initially in *Objects*, was changed to *Objects-outside*. The test results also allowed us to identify emoji that were not placed in the most appropriate category, for example the four emoji representing *card suits* (♣♦♥♠) were changed from *Geometric shapes* to *Symbols*.

The results obtained also made us question whether new categories would be necessary. This is the case of some of the emoji from *Objects* that were marked by most participants as "none" – *tooth* 🦷 (`U+1F9B7`), *brain* 🧠 (`U+1F9E0`), *heart* 🫀 (`U+1FAC0`), *lungs* 🫁 (`U+1FAC1`), *blood* 🩸 (`U+1FA78`) and DNA 🧬 (`U+1F9EC`). Even though we understand that it can be unusual to think of these as *Objects*, creating categories for them would only be due to thematic and it would not have a significant advantage from a visual blending point of view.

On the other hand, we identified two additional categories that may facilitate the blending process. First, some emoji that were placed in *Clothing-head* due to their potential use in visual blending were labelled by most participants as belonging to the more general category *Cloth-*

ing – e.g. the *sunglasses* 🕶 (U+1F576) and the *diving mask* 🤿 (U+1F93F). Moreover, we concluded that the placement of a *crown* 👑 (U+1F451) is completely different from glasses, despite both being labelled as *Clothing-head*. As such, we decided to create the category *Clothing-Face*. Second, the coloured-heart emoji were initially placed in *Symbols-Pictorial* and most participants identified them as *Symbols*, which is more adequate as they do not resemble a physical object. In any case, the results obtained in the *Emoji Kitchen* study show that the heart emoji group have particular blending mechanisms (e.g. colour change). Moreover, there has been recent interest in the development of this emoji group – adding more heart colours.[4] We believe that these reasons are enough to justify the creation of a category for the heart emoji.

Our analysis also led to changing the description of some categories, as was the case of the description of the *People-frontal* that we mentioned earlier. Another changed description was the one of the *Faces* category as we identified that some of its emoji did not fit it. The category is described as "yellow faces showing emotions" but some emoji do not have a yellow colour, e.g. *nauseated face* 🤢 is green (U+1F922). In order to cover all its emoji, we changed the description to "smiley faces often showing emotions".

After conducting the necessary changes to the categories, we proceed to analyse the results of the concreteness task. The category with the lowest average concreteness is *Geometric Shapes* with 1.51 and the category with highest is *Animals-profile* with 4.72. We calculated the standard deviation of concreteness values of each emoji and then the average for all emoji, obtaining 0.559.

When comparing with the concreteness values used in the study described in Chapter 10, the maximum difference was in the category of *BodyPart-hair*, which had a concreteness of 5 and in the current study was assigned with 2. We calculated the average difference between concreteness values used in Chapter 10 and the ones described in the present study, obtaining a value of 0.78 with a standard deviation of 0.58, which we do not consider high values.

We produced a new version of the categorisation (Table 17.4) with updated visual concreteness values, using the results obtained. The emoji list used is aligned with *Unicode Emoji 13.1* (3577 emoji).

17.6 Emoji Kitchen Category Analysis

In Chapter 5, we presented a transformational analysis of two datasets with visual blends (*VisMet* and *Emoji Kitchen*). Most of the conclusions of the *Emoji Kitchen* analysis were case-specific, applicable to single emoji. However, we are interested in assessing if there is a way of drawing more general insight. Moreover, we identified two groups of emoji

---

4 Examining Emoji Color Spaces: A Strategy for Improving the Coverage of Heart Emoji unicode.org/L2/L2021/21075-heart-emoji-coverage.pdf

Table 17.4: Blend-oriented categorisation version 3, showing the category name, number of emoji (#E), value of visual concreteness assigned (C), example of emoji and description.

category	#E	C	e.g.	description
Activities	12	4.00	🛌	daily activities
Animals-frontal	24	4.56	🪲	frontal view of full body of animal
Animals-profile	65	4.72	🐡	profile view of full body of animal
Arrows	24	3.54	➡️	arrow like shapes
BodyPart-arms	19	4.08	💪	human limb from the shoulder to the hand
BodyPart-face	24	4.31	👁️	body parts belonging to the face
BodyPart-feet	6	4.00	🦶	terminal portion of limb from the ankle to the toes
BodyPart-fingers	6	4.00	💅	fingers but not the whole hand
BodyPart-hair	4	2.00	🦱	part of head. including hair
BodyPart-hands	181	3.77	🤌	hand doing an action or gesture
BodyPart-legs	7	4.13	🦵	human limb from the hip to the foot
Buildings	26	4.16	🛖	structure that has walls and may have a roof
Clothing	22	4.55	🧦	objects associated with clothing and accessories
Clothing-body	14	4.36	👕	wearable on the top part of the body
Clothing-Face	4	4.33	🕶️	wearable on the face
Clothing-head	8	4.19	🎩	wearable on the head
Drinks	16	4.42	🍸	drinkable
Faces	97	4.08	😋	smiley faces often showing emotions
Faces-animal	22	4.51	🐻‍❄️	show animal faces in frontal view
Faces-animal-profile	6	4.62	🦓	show animal faces in profile view
Faces-cat	9	4.26	😹	cat faces showing emotions
Faces-monkey	3	4.33	🙉	monkey faces doing gestures
Faces-other	14	3.73	🤖	face of entity, frontal view
Faces-people	222	4.28	👱	face of person, frontal view and may have accessories
Flags	264	3.21	🇵🇹	flag in frontal view without pole
Food	97	4.25	🍪	edible items
Geometric Shapes	37	1.51	◼️	geometric figure
Hearts	9	3.61	💚	coloured heart
Letters or punctuation	72	2.86	🅰️	letters or punctuation marks
Lines	4	2.00	〰️	long, narrow mark or band
Objects	307	4.06	🪠	material thing that can be seen and touched
Objects-outside	39	3.97	🚧	objects found outside
People	283	4.30	💏	people in several configurations
People-frontal	35	4.14	🙎	people, frontal view with shoulders
People-fullBody	281	4.26	🧎	full body of person
People-gesture	144	3.93	💁	people doing gestures
People-role	595	4.20	💂	people in different roles, frontal view with shoulders
Places	23	4.16	🏞️	shows a location
Plants	26	4.22	🌿	living organisms lacking the power of locomotion
Sports	256	4.56	🚴	people practicing sports
Symbols	124	2.49	♣️	signs
Symbols-Pictorial	42	3.20	🎦	signs with some pictorial features
Vehicles	56	4.31	🏎️	conveyances that transport people or goods
Weather	24	3.94	🌥️	weather related objects

("smiley faces" and "coloured hearts") that used common transformations in their blends. Based on these results, we questioned whether certain types of emoji may be prone to certain transformations. If that is the case, it could be possible to establish general patterns that can be used with the emoji dataset to produce blends.

In this section, we use the emoji categorisation to analyse the blends of the *Emoji Kitchen* dataset, with the goal of assessing the existence of transformational patterns among emoji of specific categories.

Our main question is: *can we assess which type of transformation is more suitable based on the type of objects being blended?*

17.6.1 *Experiment Setup*

As mentioned in Chapter 5, the blends from *Emoji Kitchen* are based on the *Noto Emoji* dataset. On the other hand, our categorisation was developed and tested using *Twitter*'s *Twemoji* dataset. Despite this difference, the categories still apply as most emoji are very similar, due to being regulated by *Unicode*. As such, we used the categorisation to further study *Emoji Kitchen*.

For each emoji blend, we identified the categories of the two input emoji. Upon identifying the categories of the input emoji of all blends, we reached the conclusion that the blends from *Emoji Kitchen* only use emoji from 17 out of the 45 categories from our categorisation (see Table 17.5). Moreover, the number of emoji per category is not the same in all categories, going from a minimum of one emoji (👑 in *Clothing-head*, 🙈 in *Faces-monkey* and 💯 in *Letters or punctuation*) to a maximum of 94 emoji in *Faces*.

We analysed the blends in terms of categories and transformations. The analysis is based on the taxonomy of transformations described in Chapter 5. To assess the most used transformation mechanisms for each category, we focused on the blends in which the emoji from the given category are used as modifiers (see Section 5.3.2.5 for a description of the identification of emoji as base or modifier). The reason behind this is that when the emoji is used as a base, it is often dependent on the other emoji (except for certain cases in which the base dictates the blend, e.g. the *crystal ball* 🔮 `U+1F52E`), whereas when emoji are used as modifiers, the transformations are imposed by them. As we are not interested in specific cases, focusing on the modifier seemed to be the best approach.

17.6.2 *Results*

Some categories are mostly related to specific types of blend structure (Table 17.5). This is the case of *Clothing-head* with more than 90% (149 out of 161) of the blends (with category's emoji used as modifier) being

Table 17.5: Categories used in *Emoji Kitchen*. The table shows, for each category, the total number of emoji (#E), the number of tested blends with emoji from the category (TB), tested blends in which the emoji is modifier (TM) and number of TM blends according to type blend structure: pairwise juxtaposition (PJ), replacing fusion (RF) and fusion (F).

category	emoji	# E	TB	TM	PJ	RF	F
Animals-frontal	🐙🐝🕷️🦇🦔🦠	6	63	29	10	8	11
Animals-profile	🐢🦙🦥	3	180	33	17	9	7
Astronomy	🌍🌛🌜🌟💫⭐	6	90	32	15	13	4
BodyPart-face	👁️💋	2	39	17	10	7	0
Clothing-head	👑	1	168	161	149	9	3
Faces	😈😀😁😂😃😄 😅😆😇👿😉(...)	94	1148	642	1	614	27
Faces-animal	🐨🐭🐰🐱🐵🐷 🐼🦁🦌	9	80	45	9	27	9
Faces-monkey	🙈	1	9	5	1	0	4
Faces-other	🌞🎃👻👽💀💩 🤖🤡	8	99	47	6	29	12
Food	🌭🌶️🍍🎂🥑🧀 🧁	7	217	124	13	94	17
Hearts	💙💚💛💜🖤🤍 🤎🧡❤️	9	409	343	27	1	315
Letters or punctuation	💯	1	12	3	2	1	0
Objects	🍽️🎈🎊🎗️💌💎 💐📰🔮☕	10	986	335	63	223	49
Objects-outside	🕳️☃️	2	28	5	2	0	3
Plants	🌲🌵🌷🌹🌼	5	63	20	7	7	6
Symbols	💓💔💕💖💗💘 💝💞💟♥️❣️	11	340	159	27	114	18
Weather	🌈🌪️🔥☁️	4	59	23	8	8	7

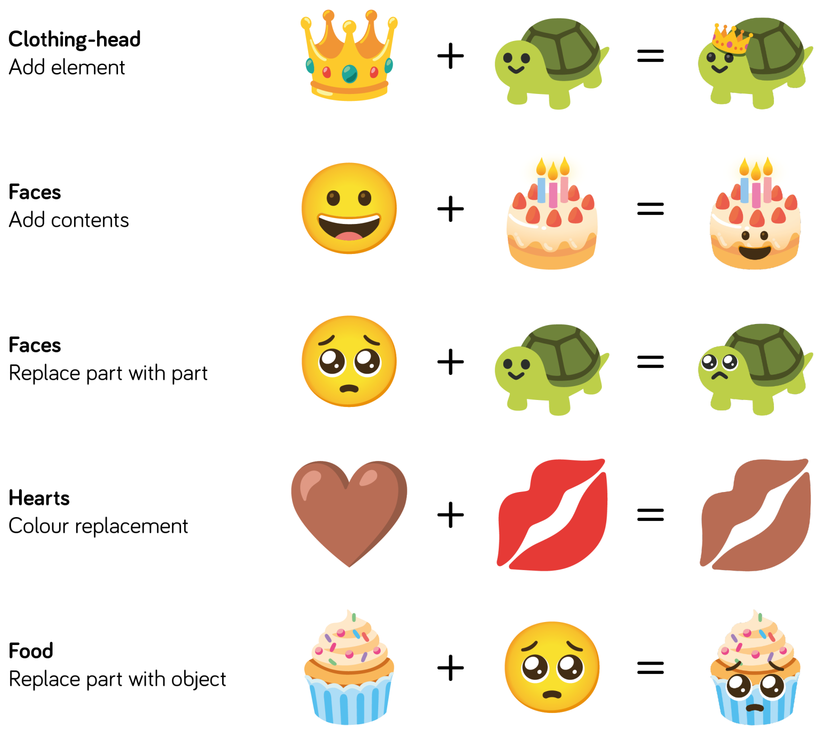


Figure 17.1: Examples of *Emoji Kitchen* blends for transformational patterns. Adapted from: *Emoji Kitchen*.

identified as Pairwise Juxtaposition (PJ), *Faces* with more than 90% as Replacing Fusion (RF) and *Hearts* with more than 90% as Fusion (F).

Regarding transformations, the most common one is "Scale", being used in more than 30% of the blends in 16 out of the 17 categories. This result is not very useful as the transformation is mostly used for positioning purposes.

For the identification of category-based transformational patterns, we focused on the ones used in more than 20 blends. As expected, some categories frequently use specific transformations (see examples in Fig. 17.1). We list the patterns below and identify the percentage of blends that use the transformation, in which the category's emoji is used as a modifier:

- *Clothing-head* uses "Add element" in more than 90% and "Rotation" in more than 75%;
- *Faces* uses "Add contents" in more than 50% and "Replace part with part" in more than 30%;
- *Hearts* uses "Colour replacement" in more than 90%;
- *Faces-animal*, *Objects* and *Symbols* use "Replace part with object" in more than 50%;
- *Food* uses "Replace part with object" in more than 75%.

The results also show that some transformations are used in most categories, such is the case of "Add element" (used in 8 out of 17), "Colour replacement" (8/17) and "Replace part with object" (9/17). "Colour replacement" is often used to improve the quality of the blend by changing the colour of the replacement part to match the one of the base emoji – for example, in *Faces-monkey* the base emoji applies its colour to the monkey arms, which are used as replacement (see Fig. 17.2). In a way, these transformations can be seen as the easy solution that can be applied in most cases.

Figure 17.2: Blend between *see-no-evil monkey* U+1F648 and *balloon* U+1F388. Source: *Emoji Kitchen*.

Our main goal was to identify which mechanisms can be used in visual blending based on the emoji category. Even though the results provide evidence that there are patterns of transformations among the emoji of each category, we are only able to study some of the categories. We are aware that by considering the prior analysis of *Emoji Kitchen* (Section 5.3.2) to improve the categorisation, these patterns were to be expected. One example of this is the creation of the *Hearts* category, which has a high rate of "Colour replacement".

In general, the results show that a category-based approach to visual blending works for some of the categories but not all. For example, the results of the *Clothing-head* category match our expectations: emoji of this category are clothing pieces to be worn on the head, thus explaining the high rate of "Add element". As already mentioned, our study has limitations as the *Emoji Kitchen* dataset only has blends using emoji from 17 out of the 45 proposed categories and for some categories only one emoji is used. Despite this, we consider that the proposed approach can be used when implementing computational systems for visual conceptual blending, providing aid in terms of transformational patterns. The advantage of using categories is that we are focused on a type of object rather than on a specific input – here "type" is to be interpreted both in terms of thematic and of style.

Nonetheless, it is important to stress that with our approach based on categories and their most common transformational patterns, we do not intend to address all these aspects but only to provide a way to generate solutions that are based on how designers produce visual blends. We do not consider this generic approach as a perfect one and we are fully aware that it will not exactly match what would be done by a designer.

17.7 SUMMARY

The main motivation behind the work described in this chapter is the development of computational systems for visual conceptual blending. Our goal is to provide a ground upon which such systems can be developed, contributing with knowledge for the development of future research on how visual blending is conducted.

In this chapter, we described the development of an emoji categorisation oriented towards visual blending. We conducted a study with users to validate the categorisation and conducted improvements based on the results obtained. Then, we described the use of the categorisation for the analysis of visual blends from the dataset *Emoji Kitchen*. Our analysis led to the identification of transformational patterns of some emoji categories.

The categorisation and identification of common transformational patterns can be considered as another emoji resource oriented towards visual blending. In any case, we believe our approach may be used with other datasets of images focused on the visual blending of single objects. However, the datasets will likely need to be pre-processed to be aligned with our categorisation, which is based on both thematic and configuration. To fully assess how our proposal can be used with other datasets, further studies are necessary.

In summary, the contributions of this chapter are: (i) the proposal of an emoji categorisation oriented towards visual blending, (ii) its evaluation through a user study, and (iii) its application in the analysis of visual blends and identification of transformational patterns.

Part VI

OUTLOOK

In the previous parts of this document, we have described the work developed within the scope of this thesis. This last part serves as an overview of the main contributions. We revisit the research questions and discuss how they have been addressed with the work that we conducted.

18 CONCLUSION

In this thesis, we investigated the use of computational approaches for the generation of visual representations of concepts. Our research hypothesis is that *computational approaches can be employed to produce visual representations of concepts, which can be useful for creativity fostering in ideation activities and for facilitating comprehension in visualisation contexts*. In addition, we identified three research questions that are related to the research hypothesis. To test the research hypothesis and address the research questions, we developed a series of systems that produce different types of visual output (pictograms, ideograms and flags), which we validated with multiple user studies.

18.1 RESEARCH QUESTIONS REVISITED

In Chapter 1, we formulated three overarching research questions. In this section, we return to these questions and discuss how the developed work has allowed us to address them.

> Research question A: *Can Computational Approaches, in particular those based on Visual Blending, be used for the visual representation of concepts?*

This research question encompasses several aspects. From a broader perspective, it concerns the capability of computational approaches to visually represent concepts. At a first stage and to be able to assess this capability, we surveyed the state of the art on computational approaches to the visual representation of concepts, which we described in Chapter 3. With our study, we described how different computational approaches can be used for the representation of different conceptual scopes – from single-word concepts to more complex structures, e.g. sentences. We have also identified the types of outputs of existing approaches, for example, some focus on icon generation while others on photorealistic compositions. Moreover, existing approaches use different representation methods – some use an inline sequential placement of images to attain a given meaning; others focus on compositional representation, in which multiple elements are brought together to produce a composition; and some employ a process of visual blending, in which elements are combined to produce a hybrid whole. Although we have explored an inline approach in the paper by Wicke and Cunha (2020), the focus of our research is blending.

When analysing the case studies presented in this thesis (*Blender*, *Emojinating* and *Moody Flags*), an important aspect to mention is that the conceptual scope is not the same in the three case studies. In the first case study (*Blender*), described in Chapter 6, we developed a system for the automatic generation of visual blends using a descriptive approach but our exploration was limited to the combination of three concepts (or *mental spaces*, as is the case): *pig*, *angel* and *cactus*. We studied the impact of *prototypical elements*, showing that representations with prototypical elements were easier to be correctly perceived, i.e. identify the two input concepts (Section 6.4). In a way, one may wonder whether the focus of this exploration is the representation of a concept or of the combination of two input concepts. The latter perspective can be seen as leading to the invention of a new concept based on a process of blending, which is aligned with the approach followed in the exploration, involving a hybrid process of blending (conceptual and visual). We consider that for visual representation the distinction between the two perspectives may be disregarded, as this approach can certainly be used to produce blends that match, and thus represent, an existing concept – e.g. *boat-house* in the experiments by Pereira and Cardoso (2002). In contrast, the *Emojinating* system produces visual representations for single-word and double-word concepts and has a broader conceptual reach, not being limited to the combination of specific concepts. Lastly, *Moody Flags* is implemented to allow the visual representation of the combination of a country, through its flag, and a given single-word topic (concept). As such, the visual representation of the concept is done by changing the flag of the country. These three approaches were used to explore different ways of visually representing concepts through computational means, thus addressing research question A. One inherent limitation of these knowledge-based approaches concerns the amount of information the systems have access to. For example, the exploration described in Section 6.4 (*Blender*) was highly limited in terms of conceptual reach, as the input data had to be manually produced for each of the three concepts (*pig*, *angel* and *cactus*), which consisted in (i) a *collection of visual representations* and their pre-processing to make them suitable to be used in the process of blending, and (ii) construction of *mental spaces* in the format of semantic networks. This sort of procedure is not scalable to be used in an open-ended system for the visual representation of concepts.

In *Moody Flags*, we also conducted a task of collection of semantic knowledge associated with flags of several countries (Section 14.3.1), as well as preparation of Scalable Vector Graphics (SVG) flag files for blending purposes. However, due to the nature of the system, this task is intrinsically limited to the number of existing flags. On the other hand, concerning the representation of the topic on the flag, the visual representation capability of the system is dependent on having a broad conceptual reach. We established the goal of supplying the system with as much knowledge as possible, giving it access to information from two additional sources beyond the data associated with the flags (Section 14.3.2): a colour name dataset and also knowledge related to emoji. Furthermore, using multiple sources of semantic information has enabled us to explore different ways of representing concepts – from low-level perceptual features (colour) to more complex structures (emoji).

For the implementation of the *Blender* and *Moody Flags*, we conducted tasks related to the preparation of the input visual representations both in terms of structure (organising layers) and also of layer labelling. These tasks were necessary for the implemented blending process to occur and were key in the alignment between the visual and the conceptual levels. Despite this, the two tasks were the main bottleneck for the performance of the two systems. Differently, in *Emojinating* we did not conduct a process of labelling the input representations (i.e. emoji) mainly due to their high number. This consequently reduces the connection between the visual and the conceptual levels but increases the scalability of the process. Other authors employ methods of annotation to prepare the input for the process of visual blending, e.g. Chilton, Petridis, and Agrawala (2019) and Zhao et al. (2020) develop annotation interfaces. Going back to the proposal by Tamés (2009) for the development of a collaborative database that connects images and concepts, we consider that the emoji set is partially aligned with their goal through the association of different meanings to each emoji, which can be computationally explored, as we demonstrated.

In *Emojinating*, we used a method to go beyond the initial limitation of the input semantic knowledge. We developed a component called *Concept Extender* that queries *ConceptNet* to obtain concepts that are related to a given one. There are two main advantages to the use of this strategy. First, it allows the system to search for similar concepts when it does not have visual knowledge for a given word. Second, it enables the system to visually represent single-word concepts through visual blending by extending the initial concept to double-word related ones. These two points are key in achieving a broader conceptual reach. Nevertheless, this strategy also has drawbacks. By extending to related concepts, the resulting visual representation loses the connection to the initial concept. For example, we may use a *type-of* relation in *ConceptNet* to obtain a related concept. However, this does not guarantee a successful representation, as is the case if one represents a higher level category, e.g. *fruit*, by resorting to a lower-level one, e.g. *banana*. Similarly, resorting to higher-level categories to represent lower-level categories may unintentionally result in neglecting some specific attributes that are not normally shared by other categories – for example, using *flower* to represent *rose*, dismisses rose-specific properties. Despite considering *conceptual extension* strategies a sub-optimal solution, it is a way to produce visual representations in situations in which there is a lack of visual information for a given concept. Other authors have used similar strategies to obtain related concepts to a given one, e.g.

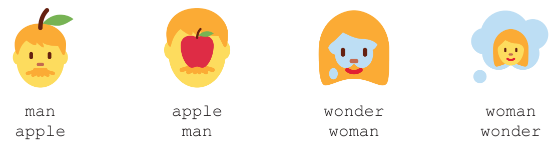


Figure 18.1: Word order in blend production

Chilton, Petridis, and Agrawala (2019) use a process of collaborative brainstorming with users. Bolognesi and Vernillo (2019) highlight the role of *metonymy* in the depiction of abstract concepts through concrete ones.

Going back to the analysis of the input of the developed systems, in the case of *Blender*, a given combination of concepts is introduced into the system, e.g. *pig-cactus*. However, the resulting blends can be considered to relate to two different views, interpreted as concepts, e.g. a *cactus pig* and a *pig cactus*. In this sense, it is more appropriate to consider that focus of the *Blender* is not exactly the representation of a specific concept, even though its results can be seen as such. A contrasting approach is used in *Emojinating* in which the input has an ordered format (the order of the words matters), e.g. *apple man* is different from *man apple* (see Fig. 18.1). In *Emojinating*, we used a modifier-head strategy, in which the first word of the concept is considered to be the modifier and the emoji used to represent it takes the role of replacement. Despite this, we consider that the connection between concept and visual representation should be further explored in order to better understand what makes the blend better represent the concept – a topic explored by Pollak et al. (2015).

In addition, even though the replacement strategy based on a modifier-head structure used in *Emojinating* works in many cases, it does not work for all: in Section 4.1.1 we have pointed out how different methods can be used for the interpretation of noun-noun compounds. This is especially complex for invented concepts – for example, *apple hat* can be a hat made of apples, a hat used in apple picking, a tiny hat made for an apple, a (probably overpriced) hat made by *Apple*, etc. From a visual blending point of view, these different interpretations may involve different blending transformations (see our proposal for a taxonomy of transformations in Section 5.2) – e.g. a representation of an *apple hat* as a "hat made of apples" may resort to the multiplication of apples positioned in a hat shape, while as "tiny hat" it would involve scaling and positioning on top of an apple. The decision between which blending strategy to use is not easy and may depend on many factors, e.g. whether the goal is to produce a faithful and realistic visual representation or a more creative and unconventional one.

Additionally, not all aspects are considered. For example, plurals are not taken into consideration in *Emojinating* and the removal of stop-words (described in Section 8.1.2) may lead to a change in meaning, e.g. "Serpent of the Year" is not the same as "Serpent Year" but the system produces the same results. Nonetheless, we do not consider these issues problematic, as they are not critical for the goals addressed in this thesis.

In regards to using blending as a method for visual representation of concepts, we conducted user studies that allowed us to investigate the subject. We analysed specific aspects related to concept representation: the impact of the concreteness of concepts in the performance of a visual blending system (Q1 in Section 10.4), the difference between representing single-word and double-word concepts (Q2 in Section 10.4) and the relation between concreteness of concepts being represented and the visual concreteness of the emoji used in their representation (Q3 in Section 10.4). In Section 11.3, we assessed the use of different blend types (*juxtaposition*, *replacement* and *fusion*) and concluded that the usefulness of *fusion* in the *Emojinating* system is somehow limited. The subject is further discussed in Section 12.3, in which a general analysis of *Emojinating* is conducted.

Our approach with blending was to explore the combination between visual blending and conceptual blending for concept representation. Visual blending on its own may not have any conceptual ground, as we have seen with the *Emoji Kitchen* dataset (Chapter 5). With its combination with a conceptual level, our goal was to produce blends that are conceptually grounded and represent the concepts on which they were based. Despite having this goal, we consider that the conceptual layer of the developed systems can be improved. In this sense, we introduce the subject of visual conceptual blending and present a roadmap for its implementation (Chapter 15).

Overall, we consider that we can answer positively to research question A. The results obtained from the studies conducted with the developed systems show that the systems can produce visual representations of concepts: in Section 6.3.1 we have shown how the blends produced by the *Blender* system for combinations of the input concepts (*pig*, *angel* and *cactus*) are similar to the blends drawn by users; a similar comparison is done between blends drawn by users and produced by *Emojinating* (Section 12.2.1.2); in Sections 8.2 and 8.3, we have described studies with blends produced by *Emojinating* that show that the system is able to visually represent concepts; and in Chapter 10, we described a study that compared the performance of *Emojinating* in the visual representation of single-word and double-word concepts, addressing topics such as the impact of concreteness. Worth of mention is an interesting result obtained in the study described in Section 12.2.1, in which users had to draw visual representations for concepts and then were shown blends produced by *Emojinating* for the same concepts. In a few cases, the user was not able to draw a representation and considered the blend a good solution. Nonetheless, we consider that the relation between blending strategy, visual representation quality and interpretation is a matter that deserves further studying.

> Research question B: *How can the user be integrated and allowed to express their preferences?*

To address this question, we employed different mechanisms with the goal of allowing the user to interact with systems and influence the production of visual representations. We explored the use of Artificial Intelligence (AI) techniques, such as Evolutionary Algorithms (EAs) and Interactive Evolutionary Computation (IEC), and different modes of interaction with the user, developing systems from fully autonomous to co-creative ones. In Chapter 9, we described the implementation of an IEC approach in *Emojinating*, in which the user is able to guide the system towards the generation of solutions with desired characteristics. The user can influence the system at two levels: at a macro level, an Estimation of Distribution Algorithm (EDA)-inspired method is used to make the system adapt to user preferences; at a micro level, the user can select individuals to be reproduced by the system. In Section 9.3 we compared the evolutionary version of the system with a deterministic one, showing that with the interaction of the user the system can evolve better solutions.

Moreover, this thesis is motivated by the possibility of developing systems that can be used to aid in ideation activities by fostering the creativity of the user. In this sense, the interaction between user and system can be explored in a way that each side influences the other, i.e. the user is able to express their preferences and affect the system but the system is also able to somehow have a sense of preference and take actions that may also have an effect on the user. This perspective is addressed in a third version of *Emojinating* (Chapter 11), in which the relation between user and system is improved to explore a more co-creative collaboration. We compared the co-creative version of the system with the evolutionary one in Section 11.3. All in all, we consider that the implemented approaches enable the user to express their preferences, which are then taken into consideration by the system for the production of new solutions, thus addressing research question B.

In questions A and B, the main concern was the ability to computationally produce visual representations. Having established that computational systems are able to produce visual representations of concepts and that the user can guide the process towards solutions that match their preference, we now focus on another important issue that is covered in research question C.

> Research question C: *How are the generated symbols perceived by users?*

To address this question, we followed a development methodology in which the implemented systems and their outputs were subjected to user testing. The research question encompasses two different perspectives, depending on who the user is: (i) user as a participant who has some kind of interaction with the system and (ii) user as an unbiased observer, who had no contact with the system.

The first perspective (user as participant) was mainly studied with *Emojinating*, as it was the system that allowed a greater interaction from the user and was subject of development based on this interaction (from evolutionary to co-creative). In this perspective, the user knows the concept and evaluates the visual representations that are shown to them. Most of the studies focused on an evaluation based on quality of representation and degree of surprise: user study with the deterministic version of *Emojinating* and double-word concepts in Section 8.3; user studies with the evolutionary version of *Emojinating* using concepts from New General Service List (NGSL) in Section 9.4, and concepts from NGSL and a double-word concept list in Section 10.4; user study with the co-creative version of *Emojinating* using double-word concepts in Section 11.3. These studies mostly focused on the assessment of the capability of the system to visually represent concepts. Nonetheless, one aspect worth mentioning is related to the user's interpretation of the system's reasoning for producing a given symbol. As the system resorts to related concepts, the produced representations are often non-literal and require some kind of interpretation or decoding from the user. Interestingly, in many cases, the user comes up with an explanation for the representation, even though it might not be consistent with the reasoning from the system. Take, for example, the representation of *freedom* that results in a symbol depicting a world map and an arrow, which can be interpreted as *freedom* related to the ability to move around. Despite this, the reasoning from the system is much more simple, having extended "freedom" to "universal right". This way, the interpretation of the user does not match the system's conceptualisation but the symbol still works as a visual representation from the user's perspective. One may say that this is a happy coincidence and an exploitation of the system, which can be framed as not properly representing the concept in a way that matches the user's interpretation. However, we argue that this coincidence perfectly aligns with the goals of the thesis in the sense that one of our objectives is to aid in ideation, which is achieved by producing a symbol that the user considers to represent the concept, even if the user's interpretation does not match the system's intention – the same can happen in human co-creation by one creator interpreting another's creation in an unintended, yet perfectly suitable, way. On the other hand, the introduction of explainability procedures may help in the understanding between system and user. We have briefly addressed the subject with *Moody Flags* by providing descriptions for the flags (Section 14.3.4). Nonetheless, explainability in creative systems is worth further studying, as highlighted by Llano et al. (2020), who provides an example of the development of an advert for a toothpaste that resorts to visual blending to produce solutions using explanations.

The second perspective (user as an unbiased observer) was studied with the three systems and involved showing the generated symbols to users and asking them for an interpretation. Regarding the first case study (*Blender*), we showed generated symbols to users and asked them to name the two input concepts (Section 6.4) – the representations with prototypical elements were easier to be correctly perceived. Then, with *Emojinating* (Section 12.2.2), we asked users to identify the concepts used to produce the visual blends. In Section 14.4, we described a user study conducted with flags produced by *Moody Flags* that had the goal of analysing the perception of the generated flags by users, who had to identify the country, the change in the flag and also the meaning of such change. These studies on perception show that the generated symbols are not always easy to be correctly perceived but can still be used for creativity fostering (leading to different creative interpretations with *Emojinating*) and visualisation purposes (e.g. triggering the interest of the observer to certain events with *Moody Flags*).

18.2 SUMMARY AND CONTRIBUTIONS

To break down the complexity of the problem addressed in this thesis, we divided the work into parts, which are reflected in the different parts of this document. In the following paragraphs, we present an overview of the several parts of this thesis.

Part I: State of the Art *· Visual Representation of Concepts · Computational Approaches*

In Part I, we provided an overview of core concepts related to the work described in this thesis and we presented existing approaches to the visual representation of concepts through computational means. First, in Chapter 2, we introduce the reader to the visual representation of concepts, addressing the definition of *concept*, *visual properties*, *semiotics* and *visual grammar*. Then, in Chapter 3 we started by introducing the areas of *Computational Design*, *Computational Creativity* and *Computational Co-creativity*, and then we reviewed the state of the art in terms of computational approaches to the production of visual representations of concepts.

Part II: Intro to Blending *· Conceptual and Visual Blending*

In Part II, we presented *Blending* as an approach to the visual representation of concepts and the focus of the thesis. In Chapter 4, we introduced the reader to the conceptual side of blending, i.e. *Concept Combination* and *Conceptual Blending* (Section 4.1), and to *Visual Blending* (Section 4.2), conducting an overview of both topics. We also provided an analysis of the requirements for the implementation of computational systems for the visual representation of concepts using blending (Section 4.3), establishing a bridge between the conceptual side and the visual side, and introducing the notion of Visual Conceptual Blending. In Chapter 5, we conducted a study on Visual Blending using two different image datasets (*VisMet* and *Emoji Kitchen*) to identify common transformations.

· Visual Blending Study

· Pig, Angel & Cactus Experiment

In Chapter 6, we described the development of a system for automatic generation of visual blends using a descriptive approach. The approach consisted in the development of a visual descriptive language composed of simple shapes, attribute-based and positioning relations, bringing together the conceptual and visual aspects. We conducted experiments using three base concepts: *pig*, *angel* and *cactus*. The system (*Blender*) was composed of two modules: (i) the *Mapper* – receives the input spaces of two concepts and produces analogies; (ii) the *Visual Blender* – produces visual blends using the mappings produced by the *Mapper*, visual representations for the two concepts and corresponding list of relations among representation parts. The experimental results showed that the *Blender* is able to create analogies from input mental spaces and produce blends that follow the rules imposed by its base analogy and its relations. In our opinion, this approach allows an easier blending process and contributes to the overall sense of cohesion among the parts. The experimentation conducted in this part allowed us to identify limitations. Despite being able to take advantage of connections between the conceptual and visual side, the experiments were conducted using pre-selected and prepared input (*pig*, *angel* and *cactus*), which makes it difficult to scale to a larger number of concepts, thus being unsuitable for a multi-purpose and open-ended context. In any case, the experiments allowed us to identify Visual Blending as an approach with potential for concept visual representation.

· User study focused on perception

Our work in this part resulted in the following publications:

- João Miguel Cunha et al. (2015). "Generation of Concept-Representative Symbols." In: *Workshop Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR-WS 2015)*. CEUR
- João Miguel Cunha et al. (2017). "A Pig, an Angel and a Cactus Walk Into a Blender: A Descriptive Approach to Visual Blending." In: *Proceedings of the Eighth International Conference on Computational Creativity*

It was also supported by the following publications, which were coauthored:

- Miguel Cruz, Paul Hardman, and João Miguel Cunha (2018). "Computationally Generating Images for Music Albums." In: *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018.* Association for Computational Creativity (ACC), p. 309
- Carolina Gonçalves Lopes, João Miguel Cunha, and Pedro Martins (2020). "Towards Generative Illustration of Text." In: *Joint Proceedings of the ICCC 2020 Workshops (WS 2020)*. Ed. by Max Kreminski et al.
- Philipp Wicke and João Miguel Cunha (2020). "An Approach for Text-to-Emoji." In: *Proceedings of the Eleventh International Conference on Computational Creativity*

Part III: Emoji
· Intro to Emoji

In Part III, we explored the use of Emoji and Visual Blending for visual representation of concepts. We started by introducing the reader to emoji, highlighting their potential for the representation of concepts (Chapter 7). Then, taking into consideration the conclusions drawn in the previous part and to address the identified limitations, we decided to take advantage of the emoji connection between pictorial representation and associated semantic knowledge for the production of visual representations. This consisted in the iterative development of a system that we called *Emojinating*.

· Emojinating v1 (deterministic)

In Chapter 8, we described the first version of the *Emojinating* system, which relied on a deterministic approach and only used the emoji most semantically related to the input-concept in the blending process. *Emojinating* combines data from *ConceptNet* (Speer and Havasi, 2012), *EmojiNet* (Wijeratne et al., 2017b) and *Twitter*'s *Twemoji*. The system takes single-word or double-word concepts as input, searches existing emoji semantically related to the input concept and complements this search with a visual blending process that generates blends. There are two main tasks – retrieval of existing emoji that match the introduced concept (T1) and generation of new possibilities through visual blending (T2) – which are conducted using three components: (i) *Concept Extender* (searches *ConceptNet* for related concepts to the one introduced by the user); (ii) *Emoji Searcher* (searches emoji based on words given as input, using semantic data provided by *EmojiNet*); and (iii) *Emoji Blender* (receives two emoji as input and returns a list of possible blends). The output of the system is composed of existing emoji – i.e. emoji that directly match the query word(s) – related emoji – i.e. emoji that match a concept related to the queried one, obtained with *ConceptNet* – and generated blends. The number of elements in each of these sets varies in quantity, depending on the data found.

· Study performance with NGSL

· User study with double-word invented concepts

In order to assess the performance of the system, we used a set of 1509 nouns from the NGSL (Section 8.2) – a core vocabulary for second language learners (Browne, 2014) – and the results show that the system is able to present the user with concept-representative results for 75% of the nouns (1144 out of 1509). To further assess the system's performance, a user study was conducted with 22 participants using a list of ten randomly generated double-word concepts (Section 8.3), e.g. "Silent Snake". Blends were produced for each of the concepts using the system, which were then shown to the participants, who evaluated three aspects: ability to represent concepts, quality of the blends and degree of surprise. Overall the system was able to produce concept-representative blends and, for many cases, the participants stated that results were different from what they were expecting. Nonetheless, we identified some limitations in the conducted studies: the study with the NGSL concepts mostly focused on the capability of the system in terms of production effectiveness – i.e. whether the system is able to produce results – and on the origin of the semantic knowledge behind the emoji gathered by the *Emoji Searcher* component, giving little attention to the performance in terms of representation quality; and the study with generated concepts had little reliability due to being based on invented concepts. Moreover, the implemented system did not employ an effective strategy for exploring the search space, thus not guaranteeing that the obtained solutions were the best.

· Emojinating v2 (evolutionary)

In order to improve the exploration of the search space and allow the user to be able to influence the results of the system, we developed an interactive evolutionary approach, which we described in Chapter 9. The approach has a two-level evolution: on a macro level, it uses a method that takes inspiration from Estimation of Distribution Algorithms to direct the search to areas that match the user preference; on a micro and more specific level, it uses a standard Evolutionary Algorithm to focus the evolution on certain individuals. With the evolutionary approach, the user is able to assign fitness to individuals and store them in the archive.

· User study deterministic vs evolutionary

To assess the performance of the evolutionary system, we conducted two user studies. First, we compared the two versions of the system (*deterministic* and *evolutionary*) using a user survey with 31 participants (Section 9.3). The results showed that the evolutionary approach was able to produce improved blends in four out of the ten concepts. Second, we conducted a survey with eight participants, in which each participant generated blends for nine concepts from the NGSL (Section 9.4). The results showed that the evolutionary approach allows the exploration of more of the search space and is able to present the user with better solutions.

· User study with NGSL

· User study single-word vs double-word

Even though we had already conducted several studies, we considered that the number of analysed concepts was still low and, aside from the randomly generated concepts (Section 8.3), only single-word concepts had been tested. Moreover, the analysis in Chapter 9 mostly focused on the performance of the evolutionary approach, overlooking questions regarding the blends (e.g. emoji used), and no statistical analysis was conducted. For these reasons, we tested the system with a double-word concept list, conducted further user-testing with the NGSL dataset (single-word concepts), and compared the results, which we describe in Chapter 10. The study used a mixed-methods approach, combining quantitative (e.g. number of generations, evaluated blends per generation, etc.) and qualitative methods (how well a concept is represented by the system and surprise degree), and compared the performance of the system with single-word concepts and double-word concepts, focusing on output quality and impact of concreteness in the blending process. We identified that participants were exploiting the system by selecting blends in which one of the emoji was hidden – this mostly occurs in concrete single-word concepts. Overall, the results obtained indicate that our visual blending approach is less useful in the representation of single-word concepts.

· Emojinating v3 (co-creative)

Despite being able to evolve solutions that match the user's preferences, we considered the system to have a somewhat passive behaviour, as the actions of the system are mostly directly triggered by the user. In Chapter 11, we described a new version of *Emojinating*, which was developed with the goal of instilling a more active behaviour to the system, allowing it to self-evaluate and adapt to context, and increasing the capability of adequately responding to user actions, thus improving the co-creative relation. This way, we enhance the creative behaviour of the system, increasing its autonomy and improving the cooperative character of its interaction with the user. In general, the system is able to learn from the user behaviour, which is observed in the storing of similar blends in its archive (e.g. if the user selects blends with a large exchanged area, the system tends to replace the blends in its archive to match the user preference). Moreover, the system archive is useful to highlight blends that the user may have missed. The interaction with the system allows the user to evolve solutions that match their preferences and, at the same time, both the user and the system influence the perception of one another, leading to novel ideas.

· User study evolutionary vs co-creative

Regarding visual blend types, the deterministic and evolutionary versions of the system only produce blends using *juxtaposition* and *replacement*. In the co-creative version, we implemented *fusion*. In Section 11.3, we compare the impact of different types of blend in the generation process by conducting a user study. The results show that juxtaposition and replacement are used in the majority of the exported blends. Fusion does not seem to be very useful and may only add unnecessary variability. Nonetheless, the different blend types may have different advantages and, for this reason, *fusion* may be useful for specific kinds of blends (e.g. involving faces).

· User study on usefulness and perception

In Chapter 12, we made an overall analysis of the development of *Emojinating*. Then, we conducted a user study focused on the analysis of two topics related to the production of visual blends: usefulness and perception. The results showed the benefits of using *Emojinating* in ideation and for creativity fostering.

Our work in this part resulted in the following publications:

- João Miguel Cunha, Pedro Martins, and Penousal Machado (2018b). "How Shell and Horn make a Unicorn: Experimenting with Visual Blending in Emoji." In: *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018.* Pp. 145–152

- João Miguel Cunha, Pedro Martins, and Penousal Machado (2018a). "Emojinating: Representing Concepts Using Emoji." In: *Workshop Proceedings of the Twenty-Sixth International Conference on Case-Based Reasoning (ICCBR 2018), Stockholm, Sweden*, p. 185

- João M. Cunha et al. (2019a). "Emojinating: Evolving Emoji Blends." In: *Proceedings of the Eighth International Conference on*

Computational Intelligence in Music, Sound, Art and Design, EvoMUSART 2019, Held as Part of EvoStar 2019, Leipzig, Germany, April 24-26, 2019. Ed. by Anikó Ekárt, Antonios Liapis, and María Luz Castro Pena. Cham: Springer International Publishing, pp. 110–126. ISBN: 978-3-030-16667-0

- João Miguel Cunha et al. (2019b). "Assessing Usefulness of a Visual Blending System: "Pictionary Has Used Image-making New Meaning Logic for Decades. We Don't Need a Computational Platform to Explore the Blending Phenomena", Do We?" In: *Proceedings of the Tenth International Conference on Computational Creativity, UNC Charlotte, North Carolina, June 17-21, 2019.*
- João M. Cunha et al. (Sept. 2020b). "Visual Blending for Concept Representation: A Case Study on Emoji Generation." In: *New Generation Computing*. DOI: 10.1007/s00354-020-00107-x. URL: https://doi.org/10.1007/s00354-020-00107-x
- João Miguel Cunha, Pedro Martins, and Penousal Machado (2020a). "Emojinating Co-Creativity: Integrating Self-Evaluation and Context-Adaptation." In: *Proceedings of the Eleventh International Conference on Computational Creativity*

Part IV: Other Applications

In Part IV, we focused on other domains in which visual representation of concepts has application potential. In particular, we explored how the visual blending engine developed for *Emojinating* can be used for other purposes. In this part, we described its application in two different cases: *Data Visualisation* and *Flag Generation*.

· Application in Data Visualisation

In Chapter 13, we explored the use of *Emojinating*'s engine for Information Visualisation. Several authors point out that it would be advantageous for a glyph-based visualisation tool to have different types of data glyphs – graphical objects that possess visual features, which can be assigned to data variables to produce a visualisation. The tool would allow the user to choose glyphs more related to the data being represented. Such a tool is normally considered difficult to implement, as it would require a large repository of glyphs prepared for data representation. We propose a strategy that uses *Emojinating*'s engine to fulfil the two requirements, allowing the production of glyphs related to the data thematic (literal and non-literal). We compare the used approaches with current glyph techniques and discuss the results.

· Moody Flags

In Chapter 14, we described the development of a system called *Moody Flags*, which generates flags based on trending topics of countries, retrieved from real-time news. The process of producing flags involves the search for elements that match a queried word, which are then used to transform the original flag. The first step was the construction of a dataset of semantic and visual flag data. The dataset is used for the queried word search, which is conducted in three different places: elements of existing flags, colour names and emoji (using *Emojinating*'s engine). We explored the notion of "mutable flag" and combined it with the concept of reactivity. Our main goal was to generate flags that continuously changed to represent current events, instilling a quality of "being alive" into them and assessing how visual elements of a flag can be used to encode concepts in their design.

· User study on perception

To assess the perception of generated flags, we conducted a user study with 16 participants. The results show that the participants can identify the original flag but they have difficulty in identifying the meaning of the changes applied to the flags. The user study points to the potential of generated flags to raise awareness of certain events. Moreover, we aimed to explore the limits of the use of a national flag, questioning topics of flag misappropriation and national identity.

Our work in this part resulted in the following publications:

- João Miguel Cunha et al. (2018). "The Many-Faced Plot: Strategy for Automatic Glyph Generation." In: *Proceedings of the 22st International Conference Information Visualisation (IV), 2018*. IEEE Computer Society
- João Miguel Cunha et al. (2020c). "Ever-changing Flags: Trend-driven Symbols of Identity." In: *8th Conference on Computation, Communication, Aesthetics & X (xCoAx 2020)*. Ed. by Mario Verdicchio et al.
- João Miguel Cunha, Pedro Martins, and Penousal Machado (2020b). "Ever-changing Flags: Impact and Ethics of Modifying National Symbols." In: *Proceedings of the Eleventh International Conference on Computational Creativity*

We also presented our work at the European Conference on Artificial Intelligence 2020: "Flags of Change: Representing the Mood of a Country" ECAI 2020 (ART AND ARTIFICIAL INTELLIGENCE Workshop).

Part V: Towards Visual Conceptual Blending (VCB)

· Roadmap for VCB

· Image Schemas

· Categorisation oriented towards visual blending

In Part V, we focused on open questions of visual representation of concepts, proposing a roadmap on future developments that can lead to the implementation of visual conceptual blending systems (Chapter 15) and describing useful resources. Despite having developed techniques for the visual representation of concepts and demonstrated their potential through a series of user studies, there are several domains that can be explored to improve existing systems. In Chapter 16, we propose an approach to include affordance-related features in systems for the visual representation of concepts, by using image schemas. We first deconstruct often used icons to show the role of image schemas, then we use examples to illustrate how visual representations can be produced using image schemas. In Chapter 17, we presented the development of a categorisation of emoji oriented towards visual blending. The categorisation is based on three criteria: (i) distinction between entities, objects and places; (ii) thematic (e.g. *Animals* and *Food*); and (iii) visual characteristics and similarity (e.g. distinguishing between faces and full bodies). Our goal was to produce a useful resource for visual conceptual blending, taking advantage of the emoji set as a coherent image dataset, as well as, its large associated semantic knowledge. We validate the categorisation through a user study conducted on the crowdsourcing platform *Appen*. In the last part of the chapter, we use the categorisation to conduct a second analysis of the images of *Emoji Kitchen*. This analysis allowed us to identify transformational patterns for specific emoji categories. With this work, we aimed at providing support for future research on visual conceptual blending.

· Validation with User study

· Emoji Kitchen analysis with the categorisation

Our work in this part resulted in the following publications:

- João Miguel Cunha, Pedro Martins, and Penousal Machado (2018c). "Using Image Schemas in the Visual Representation of Concepts." In: *Joint Proceedings of the Workshops C3GI: The 7th International Workshop on Computational Creativity, Concept Invention, and General Intelligence ISD4: The 4th Image Schema Day, and SCORE: From Image Schemas to Cognitive Robotics, Bozen-Bolzano, Italy, December 13-15, 2018.* CEUR
- João Miguel Cunha, Pedro Martins, and Penousal Machado (2020d). "Let's Figure This Out: A Roadmap for Visual Conceptual Blending." In: *Proceedings of the Eleventh International Conference on Computational Creativity*
- João Miguel Cunha, Pedro Martins, and Penousal Machado (2020c). "Knowledge in Computational Design: Typography, Emoji and Flags." In: *Joint Proceedings of the ICCC 2020 Workshops (WS 2020)*. Ed. by Max Kreminski et al.

Concerning the dissemination of our research, most of the contributions were presented at international conferences and published in international journals, and have been listed throughout this section. We were able to disseminate our work in several areas, such as Computational Creativity (e.g. Cunha et al., 2017), Information Visualisation (e.g. Cunha et al., 2018) and Computational Design (e.g. Cunha, Martins, and Machado, 2020c).

Our work has won the Best Poster Award at EvoStar 2019 – poster "Emojinating: Evolving Emoji Blends" by João Miguel Cunha, Nuno Lourenço, João Correia, Pedro Martins and Penousal Machado, based on the work described by Cunha et al. (2019a).

We were invited to attend the Dagstuhl Seminar "Computational Creativity Meets Digital Literary Studies" to which we contributed with a presentation. The abstract of the presentation was published in the Dagstuhl Reports:

- João Miguel Cunha and Amílcar Cardoso (2019). "From Conceptual Blending to Visual Blending And Back." In: *Computational Creativity Meets Digital Literary Studies (Dagstuhl Seminar 19172)*

– *Dagstuhl Reports*. Ed. by Tarek Richard Besold et al. Vol. 9. 4, p. 92

In addition, we disseminated our work on exhibitions and demonstration sessions.

- "Emojinating: Hooked Beings" was part of the exhibition of Artech 2019 (Cunha, Martins, and Machado, 2019);
- "Emojinating: Hooked Beings" was exhibited at *GO Romaria Cultural*, Gouveia, July 2021;
- "Ever-changing Flags" was part of the demo session at ICCC'20 (Cunha, Martins, and Machado, 2020b).

We were invited to integrate the team of columnists from "The Creativity Post" to write about the development of our projects. Despite having a blog format, we have taken advantage of "The Creativity Post" to increase our reach in regards to the general public.

In this thesis, we developed several explorations to address the computational generation of visual representations of concepts, which we evaluated with user studies focused on effectiveness and perception. We have shown that Computational systems for the visual representation of concepts can be used to aid in ideation processes by stimulating creativity. As for future work, we consider that our research has opened paths for future research. First, the work developed on visual blending and its transformational mechanisms still has many aspects to be explored. Second, we have explored the production of concept-representative symbols but we have not addressed how these can be explored in a language system. In this sense, a possible line of research is to study how visual languages can be produced through computational means. Third, we view our proposal of a roadmap for visual conceptual blending as one of the first steps in a line of research that we hope will lead systems with a better integration of the conceptual and visual levels.

In parallel and, in great part, due to the work described in this thesis, we had the honour to participate in the organisation of several scientific activities, from which we highlight the International Conference on Computational Creativity (ICCC), and to be invited to take a role in the board of international associations, such as the *Association for Computational Creativity*. We also had the pleasure to integrate the *Computational Creativity Task Force* (Cunha et al., 2020a), a working group established to support the collective advancement of the Computational Creativity (CC) research community, which allowed us to collaborate with other researchers, who shared our the passion towards CC (a brief mention to Anna Kantosalo, Christian Guckelsberger, Kazjon Grace, Paul M. Bodily and Sarah Harmon). These activities have allowed us to broaden our perspective on the field of CC and to take part in its development, which we consider a major contribution of this thesis.

BIBLIOGRAPHY

Abbing, Roel Roscam, Peggy Pierrot, and Femke Snelting (2017). "Modifying the universal." In: *Executing Practices*, p. 33 (cit. on p. 152).

Aberman, Kfir, Jing Liao, Mingyi Shi, Dani Lischinski, Baoquan Chen, and Daniel Cohen-Or (2018). "Neural best-buddies: sparse cross-domain correspondence." In: *ACM Transactions on Graphics (TOG)* 37.4, p. 69 (cit. on p. 59).

Adobe (2019). *Emoji Trend Report 2019*. https://www.slideshare.net/adobe/adobe-emoji-trend-report-2019/1. [Online; accessed May 2020] (cit. on pp. 143, 155).

Aggarwal, Gunjan and Devi Parikh (2020). "Neuro-Symbolic Generative Art: A Preliminary Study." In: *Proceedings of the 11th International Conference on Computational Creativity*. Ed. by Tony Veale F. Amílcar Cardoso Penousal Machado and João Miguel Cunha. Coimbra, Portugal: Association for Computational Creativity, pp. 492–495. ISBN: 978-989-54160-2-8. URL: http://computationalcreativity.net/iccc20/papers/136-iccc20.pdf (cit. on p. 70).

Agkathidis, Asterios (2015). *Generative design: Form-finding Techniques in Architecture*. Laurance King Publishing (cit. on pp. 42, 43).

Agnese, Jorge, Jonathan Herrera, Haicheng Tao, and Xingquan Zhu (2020). "A survey and taxonomy of adversarial neural networks for text-to-image synthesis." In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10.4, e1345 (cit. on p. 68).

Ai, Wei, Xuan Lu, Xuanzhe Liu, Ning Wang, Gang Huang, and Qiaozhu Mei (2017). "Untangling Emoji Popularity Through Semantic Embeddings." In: *ICWSM*, pp. 2–11 (cit. on p. 145).

Anderson, Edgar (1957). "A semigraphical method for the analysis of complex problems." In: *Proc. of the National Academy of Sciences* 43.10, pp. 923–927 (cit. on p. 241).

Arnheim, Rudolf (1974a). *Art and visual perception: A psychology of the creative eye*. Univ of California Press (cit. on pp. 14, 29).

– (1974b). *Entropy and art: An essay on disorder and order*. Univ of California Press (cit. on p. 17).

Audry, Sofian and Jon Ippolito (2019). "Can Artificial Intelligence Make Art without Artists? Ask the Viewer." In: *Arts*. Vol. 8. 1. Multidisciplinary Digital Publishing Institute, p. 35 (cit. on p. 2).

Ayzenshtadt, Viktor, Christoph Langenhan, Saqib Bukhari, Klaus-Dieter Althoff, Frank Petzold, and Andreas Dengel (2017). "Extending the flexibility of case-based design support tools: A use case in the architectural domain." In: *International Conference on Case-Based Reasoning*. Springer, pp. 46–60 (cit. on p. 41).

Bäck, Thomas, Frank Hoffmeister, and Hans-Paul Schwefel (1991). "A Survey of Evolution Strategies." In: *Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA, July 1991*. Ed. by Richard K. Belew and Lashon B. Booker. Morgan Kaufmann, pp. 2–9 (cit. on p. 206).

Bai, Qiyu, Qi Dan, Zhe Mu, and Maokun Yang (2019). "A systematic review of emoji: Current research and future perspectives." In: *Frontiers in psychology* 10, p. 2221 (cit. on p. 300).

Barbieri, Francesco, Luis Espinosa-Anke, Miguel Ballesteros, Horacio Saggion, et al. (2017). "Towards the understanding of gaming audiences by modeling Twitch emotes." In: *Third Workshop on Noisy User-generated Text (W-NUT 2017); 2017 Sep 7; Copenhagen, Denmark. Stroudsburg (PA): ACL; 2017. p. 11-20.* ACL (Association for Computational Linguistics) (cit. on p. 148).

Barbieri, Francesco, Francesco Ronzano, and Horacio Saggion (2016). "What does this Emoji Mean? A Vector Space Skip-Gram Model for Twitter Emojis." In: *LREC* (cit. on pp. 145, 300, 301).

Barsalou, Lawrence W (1983). "Ad hoc categories." In: *Memory & cognition* 11.3, pp. 211–227 (cit. on p. 15).

Bateman, John A, Joana Hois, Robert Ross, and Thora Tenbrink (2010). "A linguistic ontology of space for natural language processing." In: *Artificial Intelligence* 174.14, pp. 1027–1071 (cit. on pp. 290, 292).

Bateman, John, Janina Wildfeuer, and Tuomo Hiippala (2017). *Multimodality: Foundations, research and analysis–a problem-oriented introduction*. Walter de Gruyter GmbH & Co KG (cit. on p. 292).

Bau, David, Steven Liu, Tongzhou Wang, Jun-Yan Zhu, and Antonio Torralba (2020). "Rewriting a deep generative model." In: *European Conference on Computer Vision*. Springer, pp. 351–369 (cit. on p. 67).

Becker, Julia C et al. (2017). "What do national flags stand for? An exploration of associations across 11 countries." In: *Journal of Cross-Cultural Psychology* 48.3, pp. 335–352 (cit. on p. 266).

Bennett, Brandon and Claudia Cialone (2014). "Corpus Guided Sense Cluster Analysis: a methodology for ontology development (with examples from the spatial domain)." In: *FOIS*, pp. 213–226 (cit. on pp. 287, 292).

Bentley, Peter J and David W Corne (2002). *Creative evolutionary systems*. Morgan Kaufmann (cit. on p. 44).

Berov, Leonid and Kai-Uwe Kuhnberger (2016). "Visual Hallucination For Computational Creation." In: *Proceedings of the Seventh International Conference on Computational Creativity* (cit. on p. 63).

Bertin, Jacques (2011). *Semiology of Graphics. Diagrams, networks, maps. Redlands*. Redlands California, Esri Press (cit. on pp. 11, 13, 31).

Besold, Tarek R, Maria M Hedblom, and Oliver Kutz (2017). "A narrative in three acts: Using combinations of image schemas to model events." In: *Biologically inspired cognitive architectures* 19, pp. 10–20 (cit. on p. 288).

Biederman, Irving (1987). "Recognition-by-components: a theory of human image understanding." In: *Psychological review* 94.2, p. 115 (cit. on p. 17).

Black, Alison, Paul Luna, Ole Lund, and Sue Walker (2017). *Information design: research and practice*. Taylor & Francis (cit. on p. 292).

Black, Max (1962). *Models and Metaphors*. Ithaca: Cornel Unversity Press (cit. on p. 20).

Bliss, Charles K (1965). *Semantography (Blissymbolics): A Logical Writing for an illogical World*. Semantography Blissymbolics Publ (cit. on pp. 23, 24, 37).

Boden, Margaret A (1998). "Creativity and artificial intelligence." In: *Artificial intelligence* 103.1-2, pp. 347–356 (cit. on pp. 45, 73).

– (2004). *The Creative Mind: Myths and Mechanisms*. Routledge (cit. on pp. 45, 283).

Bodily, Paul Mark (2018). "Machine Learning for Inspired, Structured, Lyrical Music Composition." PhD thesis (cit. on p. 45).

Bolognesi, Marianna (2017). "Conceptual Metaphors and Metaphoric Expressions in Images." In: *Cognitive Modelling in Language and Discourse across Cultures*. Ed. by Annalisa Baicchi and Erica Pinelli. Cambridge Scholars Publishing. Chap. XXIII (cit. on p. 82).

Bolognesi, Marianna, Romy van den Heerik, and Esther van den Berg (2018). "VisMet 1.0: An online corpus of visual metaphors." In: *Visual Metaphor*. Ed. by Gerard J. Steen. John Benjamins Publishing Company. Chap. 4 (cit. on pp. 81, 99, 100, 275).

Bolognesi, Marianna and Paola Vernillo (2019). "How abstract concepts emerge from metaphorical images: The metonymic way." In: *Language & Communication* 69, pp. 26–41 (cit. on pp. 20, 324).

Borgo, Rita, Johannes Kehrer, David HS Chung, Eamonn Maguire, Robert S Laramee, Helwig Hauser, Matthew Ward, and Min Chen (2013). "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications." In: *Eurographics (STARs)*, pp. 39–63 (cit. on p. 241).

Breault, Vincent, Sébastien Ouellet, Sterling Somers, and Jim Davies (2013). "SOILIE: A computational model of 2D visual imagination." In: *Proceedings of the 11th International Conference on Cognitive Modeling, Ottawa: Carleton University* (cit. on p. 57).

Brock, Andrew, Jeff Donahue, and Karen Simonyan (2018). "Large scale GAN training for high fidelity natural image synthesis." In: *arXiv preprint arXiv:1809.11096* (cit. on p. 64).

Brown, RL (1985). "Methods for the graphic representation of systems simulated data." In: *Ergonomics* 28.10 (cit. on p. 249).

Browne, Charles (2014). "A new general service list: The better mousetrap we've been looking for." In: *Vocabulary Learning and Instruction* 3.1, pp. 1–10 (cit. on pp. 166, 177, 184, 191, 194, 197, 330).

Bruce, Vicki, Patrick R Green, and Mark A Georgeson (2003). *Visual perception: Physiology, psychology, & ecology*. Psychology Press (cit. on p. 12).

Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (2014). "Concreteness ratings for 40 thousand generally known English word lemmas." In: *Behavior research methods* 46.3, pp. 904–911 (cit. on pp. 185, 194, 215).

Bui, Duy, Carlos Nakamura, Bruce E Bray, and Qing Zeng-Treitler (2012). "Automated illustration of patients instructions." In: *AMIA Annual Symposium Proceedings*. Vol. 2012. American Medical Informatics Association, p. 1158 (cit. on p. 55).

Burge, Jeremy (2013). *Emojipedia* (cit. on p. 300).

Burt, Peter J and Edward H Adelson (1983). "A multiresolution spline with application to image mosaics." In: *ACM Transactions on Graphics (TOG)* 2.4, pp. 217–236 (cit. on p. 81).

Burton, Ed (1995). "Thoughtful drawings: A computational model of the cognitive nature of children's drawing." In: *Computer Graphics Forum*. Vol. 14. 3. Wiley Online Library, pp. 159–170 (cit. on p. 49).

Calude, Cristian S and John P Lewis (2012). "Is there a universal image generator?" In: *Applied Mathematics and Computation* 218.16, pp. 8151–8159 (cit. on p. 2).

Cappallo, Spencer, Thomas Mensink, and Cees GM Snoek (2015). "Image2emoji: Zero-shot emoji prediction for visual media." In: *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, pp. 1311–1314 (cit. on pp. 300, 301).

Cardoso, Amílcar, Pedro Martins, Filipe Assunção, João Correia, and Penousal Machado (2015). "A Distributed Approach to Computational Creativity." In: *Intelligent Distributed Computing IX - Proceedings of the 9th International Symposium on Intelligent Distributed Computing - IDC'2015, Guimarães, Portugal, October 2015*, pp. 3–12. DOI: 10.1007/978-3-319-25017-5_1. URL: https://doi.org/10.1007/978-3-319-25017-5_1 (cit. on p. 285).

Carroll, Noel (1994). "Visual metaphor." In: *Aspects of metaphor*. Springer, pp. 189–218 (cit. on pp. 1, 82).

Casey, Elizabeth J (1987). "Visual display representation of multidimensional systems: The effect of system structure: And display integrality." In: *Proc. of the Human Factors Society Annual Meeting*. Vol. 31. 1. SAGE Publications (cit. on p. 249).

Casner, Stephen M (1991). "Task-analytic approach to the automated design of graphic presentations." In: *ACM Trans. on Graphics (ToG)* 10.2, pp. 111–151 (cit. on p. 242).

Cavazzana, Alessandro and Marianna Bolognesi (2020). "Uncanny resemblance: Words, pictures, and conceptual representations in the field of metaphor." In: *Cognitive Linguistic Studies* 7.1, pp. 31–57 (cit. on pp. 82, 83, 86–88).

Chan, Joel, Pao Siangliulue, Denisa Qori McDonald, Ruixue Liu, Reza Moradinezhad, Safa Aman, Erin T Solovey, Krzysztof Z Gajos, and Steven P Dow (2017). "Semantically far inspirations considered harmful? accounting for cognitive states in collaborative ideation." In: *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pp. 93–105 (cit. on pp. 2, 74).

Chan, Wing-Yi, Huamin Qu, and Wai-Ho Mak (2010). "Visualizing the semantic structure in classical music works." In: *IEEE Trans. on visualization and computer graphics* 16.1 (cit. on p. 249).

Chandler, Daniel (2007). *Semiotics: the basics*. Routledge (cit. on pp. 23–25).

Chau, Michael (2011). "Visualizing web search results using glyphs: Design and evaluation of a flower metaphor." In: *ACM Trans. on Management Information Systems* 2.1 (cit. on pp. 246, 248, 249).

Chen, Liuqing (2019). "Data-driven and Machine Learning based Design Creativity." PhD thesis. Dyson School of Design Engineering, Imperial College London (cit. on pp. 67, 274).

Chen, Liuqing, Pan Wang, Hao Dong, Feng Shi, Ji Han, Yike Guo, Peter RN Childs, Jun Xiao, and Chao Wu (2019a). "An artificial intelligence based data-driven approach for design ideation." In: *Journal of Visual Communication and Image Representation* 61, pp. 10–22 (cit. on p. 67).

Chen, Yifu, Zongsheng Wang, Bowen Wu, Mengyuan Li, Huan Zhang, Lin Ma, Feng Liu, Qihang Feng, and Baoxun Wang (2019b). "MemeFaceGenerator: Adversarial Synthesis of Chinese Meme-face from Natural Sentences." In: *arXiv preprint arXiv:1908.05138* (cit. on p. 68).

Chernoff, Herman (1973). "The use of faces to represent points in k-dimensional space graphically." In: *Journal of the American Statistical Association* 68.342, pp. 361–368 (cit. on pp. 241, 242).

Chilton, Lydia B, Ecenaz Jen Ozmen, Sam H Ross, and Vivian Liu (2021). "VisiFit: Structuring Iterative Improvement for Novice Designers." In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14 (cit. on pp. 94–96).

Chilton, Lydia B, Ecenaz Jen Ozmen, and Sam Ross (2020). "VisiFit: AI Tools to Iteratively Improve Visual Blends." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019* (cit. on pp. 59, 60, 85–88).

Chilton, Lydia B., Savvas Petridis, and Maneesh Agrawala (2019). "VisiBlends: A Flexible Workflow for Visual Blends." In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, p. 172. DOI: 10.1145/3290605.3300402. URL: https://doi.org/10.1145/3290605.3300402 (cit. on pp. 59, 81, 85–87, 94, 116, 274, 276, 323, 324).

Chu, Hung-Kuo, Wei-Hsin Hsu, Niloy J Mitra, Daniel Cohen-Or, Tien-Tsin Wong, and Tong-Yee Lee (2010). "Camouflage images." In: *ACM Trans. Graph.* 29.4, pp. 51–1 (cit. on p. 80).

Coelho, Darius and Klaus Mueller (June 2020). "Infomages: Embedding Data into Thematic Images." In: *Computer Graphics Forum* 39.3, pp. 593–606. DOI: `10.1111/cgf.14004`. URL: `https://doi.org/10.1111/cgf.14004` (cit. on p. 54).

Cohen, Harold (1988). "How to Draw Three People in a Botanical Garden." In: *AAAI*. Vol. 89, pp. 846–855 (cit. on pp. 46, 49).

Cohn, Neil (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* A&C Black (cit. on pp. 30, 35).

Cohn, Neil, Jan Engelen, and Joost Schilperoord (2019). "The grammar of emoji? Constraints on communicative pictorial sequencing." In: *Cognitive research: principles and implications* 4.1, p. 33 (cit. on pp. 150, 151, 153).

Cohn, Neil and Tom Foulsham (Mar. 2022). "Meaning above (and in) the head: Combinatorial visual morphology from comics and emoji." In: *Memory & Cognition*. DOI: `10.3758/s13421-022-01294-2`. URL: `https://doi.org/10.3758/s13421-022-01294-2` (cit. on p. 35).

Cohn, Neil, Tim Roijackers, Robin Schaap, and Jan Engelen (2018). "Are emoji a poor substitute for words? Sentence processing with emoji substitutions." In: *Proceedings of the 40th Annual Meeting of the Cognitive Science Society, CogSci 2018, Madison, WI, USA, July 25-28, 2018*. Ed. by Chuck Kalish, Martina A. Rau, Xiaojin (Jerry) Zhu, and Timothy T. Rogers. cognitivesciencesociety.org. URL: `https://mindmodeling.org/cogsci2018/papers/0295/index.html` (cit. on p. 150).

Colton, Simon (2011). "The painting fool in new dimensions." In: *Proceedings of the 2nd International Conference on Computational Creativity*. Vol. 112 (cit. on pp. 46, 56).

– (2012). "The Painting Fool: Stories from building an automated painter." In: *Computers and Creativity*. Ed. by Jon McCormack and Mark d'Inverno. Springer (cit. on pp. 46, 47).

Colton, Simon, Amy Smith, Sebastian Berns, and Ryan Murdock (2021). "Generative Search Engines: Initial Experiments." In: *Proceedings of the 12th International Conference on Computational Creativity*. Ed. by Andrés Gómez de Silva Garza, Tony Veale, Wendy Aguilar, and Rafael Pérez y Pérez. México City, México (Virtual): Association for Computational Creativity, pp. 237–246. ISBN: 978-989-54160-3-5. URL: `https://computationalcreativity.net/iccc21/wp-content/uploads/2021/09/ICCC_2021_paper_50.pdf` (cit. on pp. 68–70).

Colton, Simon and Geraint A. Wiggins (2012). "Computational Creativity: The Final Frontier?" In: *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012*, pp. 21–26. DOI: `10.3233/978-1-61499-098-7-21`. URL: `https://doi.org/10.3233/978-1-61499-098-7-21` (cit. on p. 45).

Confalonieri, Roberto, Joseph Corneli, Alison Pease, Enric Plaza, and Marco Schorlemmer (2015). "Using argumentation to evaluate concept blends in combinatorial creativity." In: *Proceedings of the Sixth International Conference on Computational Creativity*, pp. 174–181 (cit. on pp. 58, 61, 281, 288, 293).

Cook, Michael and Simon Colton (2011). "Automated Collage Generation-With More Intent." In: *ICCC*. Citeseer, pp. 1–3 (cit. on pp. 56, 57).

Cook, Michael, Simon Colton, Alison Pease, and Maria Teresa Llano (2019). "Framing In Computational Creativity-A Survey And Taxonomy." In: *ICCC*, pp. 156–163 (cit. on pp. 259, 284).

Correia, João, Penousal Machado, Juan Romero, and Adrian Carballal (2013). "Evolving Figurative Images Using Expression-Based Evolutionary Art." In: *Iccc*, pp. 24–31 (cit. on p. 50).

Correia, João, Tiago Martins, Pedro Martins, and Penousal Machado (2016). "X-Faces: The eXploit Is Out There." In: *Proceedings of the Seventh International Conference on Computational Creativity* (cit. on pp. 59, 274).

Costello, Fintan J and Mark T Keane (2000). "Efficient creativity: Constraint-guided conceptual combination." In: *Cognitive Science* 24.2, pp. 299–349 (cit. on pp. 16, 75, 124).

– (2001). "Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.1, p. 255 (cit. on pp. 75, 124).

Costello, Fintan (2002). "Investigating creative language: People's choice of words in the production of novel noun-noun compounds." In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 24. 24 (cit. on p. 124).

Costello, Fintan and Mark T Keane (1997). "Polysemy in Conceptual Combinaton: Testing the Constraint Theory of Combination." In: *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society: August 7-10, 1997, Stanford University*. Vol. 19. Psychology Press, p. 137 (cit. on pp. 123, 124).

Coulson, Seana (2006). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction*. Cambridge University Press (cit. on p. 124).

Cramer, Henriette, Paloma de Juan, and Joel Tetreault (2016). "Sender-intended functions of emojis in US messaging." In: *Proc. of MobileHCI 2016*. ACM (cit. on p. 145).

Cruz, Miguel Machado et al. (2019). "Os olhos também ouvem: Sistema de geração de imagens de acordo com texto e som para álbuns de música." MA thesis. Universidade de Coimbra (cit. on p. 55).

Cruz, Miguel, Paul Hardman, and João Miguel Cunha (2018). "Computationally Generating Images for Music Albums." In: *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018.* Association for Computational Creativity (ACC), p. 309 (cit. on pp. 6, 91, 329).

Cui, Weiwei, Xiaoyu Zhang, Yun Wang, He Huang, Bei Chen, Lei Fang, Haidong Zhang, Jian-Guan Lou, and Dongmei Zhang (2019). "Text-to-viz: Automatic generation of infographics from proportion-related natural language statements." In: *IEEE transactions on visualization and computer graphics* 26.1, pp. 906–916 (cit. on p. 56).

Cunha, João M., Pedro Martins, and Penousal Machado (2019). "Emojinating: Hooked Beings." In: *Proceedings of the 9th International Conference on Digital and Interactive Arts*. ARTECH 2019. Braga, Portugal: Association for Computing Machinery. ISBN: 9781450372503. DOI: 10.1145/3359852.3359964. URL: https://doi.org/10.1145/3359852.3359964 (cit. on p. 336).

Cunha, João M., Sarah Harmon, Christian Guckelsberger, Anna Kantosalo, Paul M. Bodily, and Kazjon Grace (2020a). "Understanding and Strengthening the Computational Creativity Community: A Report From The Computational Creativity Task Force." In: *Proceedings of the 11th International Conference on Computational Creativity*. Ed. by F. Amílcar Cardoso, Penousal Machado, Tony Veale, and João Miguel Cunha. Coimbra, Portugal: Association for Computational Creativity, pp. 1–7. ISBN: 978-989-54160-2-8 (cit. on p. 336).

Cunha, João M., Nuno Lourenço, João Correia, Pedro Martins, and Penousal Machado (2019a). "Emojinating: Evolving Emoji Blends." In: *Proceedings of the Eighth International Conference on Computational Intelligence in Music, Sound, Art and Design, EvoMUSART 2019, Held as Part of EvoStar 2019, Leipzig, Germany, April 24-26, 2019*. Ed. by Anikó Ekárt, Antonios Liapis, and María Luz Castro Pena. Cham: Springer International Publishing, pp. 110–126. ISBN: 978-3-030-16667-0 (cit. on pp. 6, 175, 182, 184, 194, 332, 335).

Cunha, João M., Nuno Lourenço, Pedro Martins, and Penousal Machado (Sept. 2020b). "Visual Blending for Concept Representation: A Case Study on Emoji Generation." In: *New Generation Computing*. DOI: 10.1007/s00354-020-00107-x. URL: https://doi.org/10.1007/s00354-020-00107-x (cit. on pp. 6, 191, 196, 197, 205, 214, 223, 224, 230, 299, 333).

Cunha, João Miguel and Amílcar Cardoso (2019). "From Conceptual Blending to Visual Blending And Back." In: *Computational Creativity Meets Digital Literary Studies (Dagstuhl Seminar 19172) – Dagstuhl Reports*. Ed. by Tarek Richard Besold, Pablo Gervás, Evlyn Gius, and Sara Schulz. Vol. 9. 4, p. 92 (cit. on pp. 6, 273, 335).

Cunha, João Miguel, João Gonçalves, Pedro Martins, Penousal Machado, and Amílcar Cardoso (2017). "A Pig, an Angel and a Cactus Walk Into a Blender: A Descriptive Approach to Visual Blending." In: *Proceedings of the Eighth International Conference on Computational Creativity* (cit. on pp. 5, 274, 329, 335).

Cunha, João Miguel, Pedro Martins, Amílcar Cardoso, and Penousal Machado (2015). "Generation of Concept-Representative Symbols."

In: *Workshop Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR-WS 2015)*. CEUR (cit. on pp. 2, 5, 329).

Cunha, João Miguel, Pedro Martins, and Penousal Machado (2018a). "Emojinating: Representing Concepts Using Emoji." In: *Workshop Proceedings of the Twenty-Sixth International Conference on Case-Based Reasoning (ICCBR 2018), Stockholm, Sweden*, p. 185 (cit. on pp. 6, 157, 166, 194, 332).

– (2018b). "How Shell and Horn make a Unicorn: Experimenting with Visual Blending in Emoji." In: *Proceedings of the Ninth International Conference on Computational Creativity, Salamanca, Spain, June 25-29, 2018.* Pp. 145–152 (cit. on pp. 6, 157, 171, 274, 332).

Cunha, João Miguel, Pedro Martins, and Penousal Machado (2018c). "Using Image Schemas in the Visual Representation of Concepts." In: *Joint Proceedings of the Workshops C3GI: The 7th International Workshop on Computational Creativity, Concept Invention, and General Intelligence ISD4: The 4th Image Schema Day, and SCORE: From Image Schemas to Cognitive Robotics, Bozen-Bolzano, Italy, December 13-15, 2018.* CEUR (cit. on pp. 6, 287, 335).

Cunha, João Miguel, Pedro Martins, and Penousal Machado (2020a). "Emojinating Co-Creativity: Integrating Self-Evaluation and Context-Adaptation." In: *Proceedings of the Eleventh International Conference on Computational Creativity* (cit. on pp. 6, 205, 333).

– (2020b). "Ever-changing Flags: Impact and Ethics of Modifying National Symbols." In: *Proceedings of the Eleventh International Conference on Computational Creativity* (cit. on pp. 6, 251, 334, 336).

Cunha, João Miguel, Pedro Martins, and Penousal Machado (2020c). "Knowledge in Computational Design: Typography, Emoji and Flags." In: *Joint Proceedings of the ICCC 2020 Workshops (WS 2020)*. Ed. by Max Kreminski, Viktor Eisenstadt, Sofia Pinto, and Oliver Kutz (cit. on pp. 6, 335).

– (2020d). "Let's Figure This Out: A Roadmap for Visual Conceptual Blending." In: *Proceedings of the Eleventh International Conference on Computational Creativity* (cit. on pp. 6, 273, 335).

Cunha, João Miguel, Pedro Martins, Hugo Gonçalo Oliveira, and Penousal Machado (2020c). "Ever-changing Flags: Trend-driven Symbols of Identity." In: *8th Conference on Computation, Communication, Aesthetics & X (xCoAx 2020)*. Ed. by Mario Verdicchio, Miguel Carvalhais, Luísa Ribas, and André Rangel (cit. on pp. 6, 251, 334).

Cunha, João Miguel, Tiago Martins, Pedro Martins, João Bicker, and Penousal Machado (2016). "TypeAdviser: a type design aiding-tool." In: *Proceedings of the Workshop "Computational Creativity, Concept Invention, and General Intelligence 2016"* (cit. on p. 46).

Cunha, João Miguel, Evgheni Polisciuc, Pedro Martins, and Penousal Machado (2018). "The Many-Faced Plot: Strategy for Automatic Glyph Generation." In: *Proceedings of the 22st International Conference Infor-*

mation Visualisation (IV), 2018. IEEE Computer Society (cit. on pp. 6, 241, 334, 335).

Cunha, João Miguel, Sérgio Rebelo, Pedro Martins, and Penousal Machado (2019b). "Assessing Usefulness of a Visual Blending System: "Pictionary Has Used Image-making New Meaning Logic for Decades. We Don't Need a Computational Platform to Explore the Blending Phenomena", Do We?" In: *Proceedings of the Tenth International Conference on Computational Creativity, UNC Charlotte, North Carolina, June 17-21, 2019.* (Cit. on pp. 6, 214, 223, 224, 333).

Danesi, Marcel (2017). *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury Publishing (cit. on pp. 144, 150).

Darwin, Charles (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray (cit. on pp. 44, 175).

Davis, Nicholas Mark (2013). "Human-computer co-creativity: Blending human and computational creativity." In: *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference* (cit. on p. 47).

Davis, Nicholas, Chih-Pin Hsiao, Yanna Popova, and Brian Magerko (2015). "An enactive model of creativity for computational collaboration and co-creation." In: *Creativity in the Digital Age*. Springer, pp. 109–133 (cit. on p. 205).

Davis, Nicholas, Chih-Pin Hsiao, Kunwar Yashraj Singh, and Brian Magerko (2016). "Co-creative drawing agent with object recognition." In: *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference* (cit. on pp. 47, 65, 175).

Dawkins, Richard et al. (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. WW Norton & Company (cit. on pp. 44, 49).

De Saussure, Ferdinand (2011). *Course in general linguistics*. Columbia University Press (cit. on p. 23).

De Soete, Geert (1986). "A perceptual study of the Flury—Riedwyl faces for graphically displaying multivariate data." In: *Int. journal of man-machine studies* 25.5, pp. 549–555 (cit. on p. 249).

Dick, Philip K. (1962). *The Man in the High Castle*. Putnam (cit. on pp. 251, 253).

Dimson, Thomas (2015). *Emojineering part 1: Machine learning for emoji trends* (cit. on p. 145).

Ding, Ming, Zhuoyi Yang, Wenyi Hong, Wendi Zheng, Chang Zhou, Da Yin, Junyang Lin, Xu Zou, Zhou Shao, Hongxia Yang, et al. (2021). "CogView: Mastering Text-to-Image Generation via Transformers." In: *arXiv preprint arXiv:2105.13290* (cit. on p. 69).

Dixon, Daniel, Manoj Prasad, and Tracy Hammond (2010). "icandraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 897–906 (cit. on p. 46).

Donato, Giulia and Patrizia Paggio (2017). "Investigating Redundancy in Emoji Use: Study on a Twitter Based Corpus." In: *Proc. of WASSA 2017* (cit. on pp. 145, 195, 300–302).

Dorris, Nathan, Brian Carnahan, Luke Orsini, and Lois-Ann Kuntz (2004). "Interactive evolutionary design of anthropomorphic symbols." In: *Evolutionary Computation, 2004. CEC2004. Congress on*. Vol. 1. IEEE, pp. 433–440 (cit. on pp. 51, 176).

Dozier, Gerry, Brian Carnahan, Cheryl Seals, L-A Kuntz, and Ser-Geon Fu (2005). "An interactive distributed evolutionary algorithm (IDEA) for design." In: *Systems, Man and Cybernetics, 2005 IEEE Int. Conf. on*. Vol. 1. IEEE, pp. 418–422 (cit. on pp. 51, 176).

Duarte, José Pinto (2005). "Towards the mass customization of housing: the grammar of Siza's houses at Malagueira." In: *Environment and planning B: Planning and Design* 32.3, pp. 347–380 (cit. on p. 43).

Duro, Lígia, Penousal Machado, and Artur Rebelo (2012). "Graphic narratives: generative book covers." In: *International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2012, Los Angeles, CA, USA, August 5-9, 2012, Poster Proceedings*. ACM, p. 22 (cit. on p. 51).

Dürscheid, Christa and Christina Margrit Siever (2017). "Beyond the Alphabet–Communcataion of Emojis." In: *Kurzfassung eines (auf Deutsch) zur Publikation eingereichten Manuskripts* (cit. on pp. 145, 149, 152).

Eiben, Agoston E. and James E. Smith (2015). *Introduction to Evolutionary Computing*. 2nd. Springer Publishing Company, Incorporated (cit. on p. 44).

Eisner, Ben, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel (2016). "emoji2vec: Learning Emoji Representations from their Description." In: *Proc. SocialNLP* (cit. on p. 145).

Elgenius, Gabriella (2011). *Symbols of nations and nationalism: Celebrating nationhood*. Palgrave Macmillan (cit. on p. 251).

Engelhardt, Jörg von (2002). *The language of graphics: A framework for the analysis of syntax and meaning in maps, charts and diagrams*. Yuri Engelhardt (cit. on pp. 11, 13, 24–27, 29–36, 121, 122, 130, 292).

Eppe, Manfred, Ewen Maclean, Roberto Confalonieri, Oliver Kutz, Marco Schorlemmer, Enric Plaza, and Kai-Uwe Kühnberger (2018). "A computational framework for conceptual blending." In: *Artif. Intell.* 256, pp. 105–129. DOI: 10.1016/j.artint.2017.11.005. URL: https://doi.org/10.1016/j.artint.2017.11.005 (cit. on p. 279).

Evans, Vyvyan and Melanie Green (2006). *Cognitive Linguistics: An Introduction*. Edinburgh University Press (cit. on pp. 76, 78).

Falomir, Zoe, Lledó Museros, Luis Gonzalez-Abril, M Teresa Escrig, and Juan A Ortega (2012). "A model for the qualitative description of images based on visual and spatial features." In: *Computer Vision and Image Understanding* 116.6, pp. 698–714 (cit. on p. 33).

Falomir, Zoe and Enric Plaza (2019). "Towards a model of creative understanding: deconstructing and recreating conceptual blends using image schemas and qualitative spatial descriptors." In: *Annals of Mathematics and Artificial Intelligence* (cit. on p. 288).

Fan, Judith E, Monica Dinculescu, and David Ha (2019). "collabdraw: An Environment for Collaborative Sketching with an Artificial Agent." In: *Proceedings of the 2019 on Creativity and Cognition*. ACM, pp. 556–561 (cit. on p. 66).

Fares, Murhaf (2016). "A Dataset for Joint Noun-Noun Compound Bracketing and Interpretation." In: *Proceedings of the ACL 2016 Student Research Workshop*. Berlin, Germany: Association for Computational Linguistics, pp. 72–79 (cit. on pp. 194, 197, 215).

Fauconnier, Gilles (1994). *Mental Spaces: Aspects of Meaning Construction in Natural Language*. New York: Cambridge University Press (cit. on pp. 58, 76).

Fauconnier, Gilles and Mark Turner (1998). "Conceptual integration networks." In: *Cognitive science* 22.2, pp. 133–187 (cit. on pp. x, 76, 134, 135, 282).

– (2002). *The Way We Think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books (cit. on pp. 1, 58, 60, 73, 76–78).

Featherstone, Coral (2019). "Conceptual Blending Techniques for Data Visualisation." MA thesis. University of South Africa (cit. on pp. 54, 90).

Feng, Yang, Lin Ma, Wei Liu, and Jiebo Luo (2019). "Unsupervised image captioning." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4125–4134 (cit. on p. 284).

Ferreira, Diogo (2019). "Design Editorial Algorítmico." MA thesis. Universidade de Coimbra (cit. on p. 42).

Flury, Bernhard and Hans Riedwyl (1981). "Graphical representation of multivariate data by means of asymmetrical faces." In: *Journal of the American Statistical Association* 76.376 (cit. on pp. 246, 247, 249).

Forceville, Charles (1999). "Educating the eye? Kress and Van Leeuwen's reading images: The grammar of visual design (1996)." In: *Language and Literature* 8.2, pp. 163–178 (cit. on p. 30).

– (2002a). *Pictorial metaphor in advertising*. Routledge (cit. on pp. 82, 83).

– (2002b). "The identification of target and source in pictorial metaphors." In: *Journal of pragmatics* 34.1, pp. 1–14 (cit. on p. 82).

Frans, Kevin, LB Soros, and Olaf Witkowski (2021). "Clipdraw: Exploring text-to-drawing synthesis through language-image encoders." In: *arXiv preprint arXiv:2106.14843* (cit. on pp. 68, 69).

French, Robert M (2002). "The computational modeling of analogy-making." In: *Trends in cognitive Sciences* 6.5, pp. 200–205 (cit. on p. 2).

Frolov, Stanislav, Tobias Hinz, Federico Raue, Jörn Hees, and Andreas Dengel (2021). "Adversarial text-to-image synthesis: A review." In: *arXiv preprint arXiv:2101.09983* (cit. on p. 68).

Fuchs, Johannes, Petra Isenberg, Anastasia Bezerianos, and Daniel Keim (2017). "A systematic review of experimental studies on data glyphs."

In: *IEEE Trans. on Visualization and Computer Graphics* 23.7 (cit. on pp. 241, 245).

Fuchs, Johannes, Niklas Weiler, and Tobias Schreck (2015). "Leaf Glyph Visualizing Multi-Dimensional Data with Environmental Cues." In: (cit. on pp. 242, 248, 249).

Galanos, Theodoros, Antonios Liapis, and Georgios N Yannakakis (2021). "AffectGAN: Affect-Based Generative Art Driven by Semantics." In: *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, pp. 01–07 (cit. on pp. 68, 69).

Galatolo, Federico, Mario Cimino, and Gigliola Vaglini (2021). "Generating Images from Caption and Vice Versa via CLIP-Guided Generative Latent Space Search." In: *Proceedings of the International Conference on Image Processing and Vision Engineering*. DOI: 10.5220/0010503701660174 (cit. on p. 68).

Garg, Supriya, Tamara Berg, and Klaus Mueller (2011). "Iconizer: A framework to identify and create effective representations for visual information encoding." In: *International Symposium on Smart Graphics*. Springer, pp. 78–90 (cit. on pp. 53, 54).

Gatys, Leon A, Alexander S Ecker, and Matthias Bethge (2015). "A neural algorithm of artistic style." In: *arXiv preprint arXiv:1508.06576* (cit. on p. 63).

Ge-Stadnyk, Jing (2021). "Communicative functions of emoji sequences in the context of self-presentation: A comparative study of Weibo and Twitter users." In: *Discourse & Communication*, p. 17504813211002038 (cit. on p. 150).

Ge, Songwei and Devi Parikh (2021). "Visual Conceptual Blending with Large-Scale Language and Vision Models." In: *Proceedings of the 12th International Conference on Computational Creativity*. Ed. by Andr'es G'omez de Silva Garza, Tony Veale, Wendy Aguilar, and Rafael P'erez y P'erez. M'exico City, M'exico (Virtual): Association for Computational Creativity, pp. 6–10. ISBN: 978-989-54160-3-5. URL: https://computationalcreativity.net/iccc21/wp-content/uploads/2021/09/ICCC_2021_paper_90.pdf (cit. on pp. 67, 69, 70, 274, 278).

Geisler, Michael E (2005). "What Are National Symbols — and What Do They Do to Us?" In: *National symbols, fractured identities: Contesting the national narrative*. UPNE. Chap. Introduction (cit. on p. 251).

Gentner, Dedre and Michael Jeziorski (1993). "The shift from metaphor to analogy in Western science." In: *Metaphor and thought*. Ed. by Andrew Ortony. Cambridge University Press Cambridge. Chap. 20, pp. 447–480 (cit. on p. 20).

Gero, John S, Kazjon S Grace, and Robert Saunders (2008). "Computational Analogy-Making in Designning: A Process Architecture." In: (cit. on p. 41).

Gibson, James J (1977). “The theory of affordances.” In: *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Ed. by R. Shaw and J. Bransford. NJ: Lawrence Erlbaum, Hillsdale, pp. 67–82 (cit. on p. 18).

Gkiouzepas, Lampros and Margaret K Hogg (2011). “Articulating a new framework for visual metaphors in advertising.” In: *Journal of Advertising* 40.1, pp. 103–120 (cit. on p. 85).

Goel, Ashok (2019). “Computational design, analogy, and creativity.” In: *Computational Creativity*. Springer, pp. 141–158 (cit. on p. 41).

Goguen, Joseph (1999). “An Introduction to Algebraic Semiotics, with Applications to User Interface Design.” In: *Lecture Notes in Artificial Intelligence*. Vol. Computation for Metaphor, Analogy and Agents. Springer, pp. 242–291 (cit. on p. 60).

Goldberg, Andrew B, Jake Rosin, Xiaojin Zhu, and Charles R Dyer (2009). “Toward text-to-picture synthesis.” In: *NIPS 2009 Mini-Symposia on Assistive Machine Learning for People with Disabilities* (cit. on p. 55).

Goldstein, Robert Justin (2019). *Saving old glory: The history of the American flag desecration controversy*. Routledge (cit. on p. 267).

Gonçalves, João, Pedro Martins, and Amílcar Cardoso (2018). “A Fast Mapper as a Foundation for Forthcoming Conceptual Blending Experiments.” In: *Case-Based Reasoning Research and Development - 26th International Conference, ICCBR 2018, Stockholm, Sweden, July 9-12, 2018, Proceedings*. Ed. by Michael T. Cox, Peter Funk, and Shahina Begum. Vol. 11156. Lecture Notes in Computer Science. Springer, pp. 532–547. DOI: 10.1007/978-3-030-01081-2_35. URL: https://doi.org/10.1007/978-3-030-01081-2_35 (cit. on pp. 130, 208).

Gonçalves, João, Pedro Martins, António Cruz, and Amílcar Cardoso (2015). “Seeking Divisions of Domains on Semantic Networks by Evolutionary Bridging.” In: *ICCBR (Workshops)*, pp. 113–122 (cit. on p. 278).

Gonçalo Oliveira, Hugo, Diogo Costa, and Alexandre Miguel Pinto (2016). “One does not simply produce funny memes!explorations on the automatic generation of internet humor.” In: *Proceedings of Seventh International Conference on Computational Creativity* (cit. on p. 258).

Gong, Daoxiong, Jie Yan, and Guoyu Zuo (2010). “A review of gait optimization based on evolutionary computation.” In: *Applied Computational Intelligence and Soft Computing* (cit. on p. 176).

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio (2014). “Generative adversarial nets.” In: *Advances in neural information processing systems* 27 (cit. on p. 64).

Grace, Kazjon, Mary Lou Maher, David Wilson, and Nadia Najjar (2017). “Personalised specific curiosity for computational design systems.” In: *Design Computing and Cognition'16*. Springer, pp. 593–610 (cit. on p. 41).

Graves, Maitland E (1951). *Art of color and design* (cit. on p. 14).

Groot, Arjan (2000). *UNFR Index: Universal authority for National Flag Registration Index* (cit. on p. 253).

Gross, Benedikt, Hartmut Bohnacker, Julia Laub, and Claudius Lazzeroni (2018). *Generative design: Visualize, program, and create with JavaScript in p5.js*. Princeton Architectural Press (cit. on p. 43).

Guckelsberger, Christian (2020). "Intrinsic motivation in computational creativity applied to videogames." PhD thesis. Queen Mary University of London. URL: https://research.aalto.fi/en/publications/intrinsic-motivation-in-computational-creativity-applied-to-video (cit. on pp. 45, 46).

Guibon, Gaël, Magalie Ochs, and Patrice Bellot (2016). "From Emojis to Sentiment Analysis." In: *WACAI 2016* (cit. on p. 152).

Guida, Francesco E (2014). "Generative Visual Identities. New Scenarios in Corporate Identity." In: *GA2014 – XVII Generative Art Conference*. Ed. by Celestino Soddu and Enrica Colabella. Domus Argenia, pp. 121–132 (cit. on pp. 43, 252).

Gustafsson, Viktor (2017). "Replacing words with emojis and its effect on reading time." In: *USCCS 2017* (cit. on pp. 145, 150).

Ha, David and Douglas Eck (2017). "A neural representation of sketch drawings." In: *arXiv preprint arXiv:1704.03477* (cit. on pp. 65–67).

Hamner, CG, DW Turner, and DM Young (1987). "Comparisons of several graphical methods for representing multivariate data." In: *Computers & Mathematics with Applications* 13.7 (cit. on p. 249).

Hampton, James A (1987). "Inheritance of attributes in natural concept conjunctions." In: *Memory & Cognition* 15.1, pp. 55–71 (cit. on p. 76).

Hartvigsen, Kenneth (2018). "The Flag in American Art." In: *The American Flag: An Encyclopedia of the Stars and Stripes in US History, Culture, and Law*, p. 45 (cit. on p. 267).

Hassan, Enass Mahmoud Mohamed (2015). "The Semiotics of Pictogram in the Signage Systems." In: *International Design Journal* 5 (cit. on pp. 18, 29).

Hassani, Kaveh and Won-Sook Lee (2016). "Visualizing natural language descriptions: A survey." In: *ACM Computing Surveys (CSUR)* 49.1, pp. 1–34 (cit. on p. 56).

Havasi, Catherine, Robert Speer, and Justin Holmgren (2010). "Automated Color Selection Using Semantic Knowledge." In: *AAAI Fall Symposium: Commonsense Knowledge* (cit. on p. 52).

Healy, Don (1994). "Evolutionary Vexillography: One Flag's Influence in Modern Design." In: *Raven: A Journal of Vexillology* 1, pp. 41–64 (cit. on p. 251).

Heath, Derrall, David Norton, and Dan Ventura (2014). "Conveying semantics through visual metaphor." In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 5.2, pp. 1–17 (cit. on pp. 62, 63).

Heath, Derrall and Dan Ventura (2016a). "Before A Computer Can Draw, It Must First Learn To See." In: *Proceedings of the 7th Interna-*

tional Conference on Computational Creativity, page to appear (cit. on p. 63).

Heath, Derrall and Dan Ventura (2016b). "Creating Images by Learning Image Semantics Using Vector Space Models." In: (cit. on pp. 46, 62).

Hedblom, Maria M and Oliver Kutz (2015). "Shape up, baby! Perception, Image Schemas, and Shapes in Concept Formation." In: *Proceedings of the Third Interdisciplinary Workshop SHAPES 3.0 — The Shape of Things 2015* (cit. on pp. 17, 287).

Hedblom, Maria M, Oliver Kutz, Till Mossakowski, and Fabian Neuhaus (2017). "Between C ontact and S upport: Introducing a Logic for Image Schemas and Directed Movement." In: *Conference of the Italian Association for Artificial Intelligence*. Springer, pp. 256–268 (cit. on p. 288).

Hedblom, Maria M, Oliver Kutz, and Fabian Neuhaus (2015). "Choosing the right path: image schema theory as a foundation for concept invention." In: *Journal of Artificial General Intelligence* 6.1, pp. 21–54 (cit. on pp. 287, 293).

– (2016). "Image schemas in computational conceptual blending." In: *Cognitive Systems Research* 39, pp. 42–57 (cit. on pp. 287, 296).

– (2018). "Image Schemas and Concept Invention." In: *Concept Invention*. Springer, pp. 99–132 (cit. on pp. 289, 291).

Heller, Steven (2015). *American Reich*. Ed. by Design Observer. `https://designobserver.com/feature/american-reich/39109`. Accessed: January 2020 (cit. on pp. 251, 253).

Herring, Susan C and Jing Ge (2020). "Do Emoji Sequences Have a Preferred Word Order?" In: *Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media* (cit. on p. 150).

Herring, Susan and Ashley Dainas (2017). ""Nice Picture Comment!" Graphicons in Facebook Comment Threads." In: *Proc. of the 50th Hawaii Int. Conference on System Sciences* (cit. on p. 145).

Hertzmann, Aaron (2018). "Can computers create art?" In: *Arts*. Vol. 7. 2. Multidisciplinary Digital Publishing Institute, p. 18 (cit. on p. 2).

Hew, Soonhin et al. (2012). "Using combining evolution of Pictogram Chinese Characters to represent ideogrammic compounds Chinese characters." In: *Computing and Convergence Technology, 2012 7th International Conference on*. IEEE, pp. 219–223 (cit. on p. 36).

Hiroyasu, Tomoyuki, Misato Tanaka, Fuyuko Ito, and Mitsunori Miki (2008). "Discussion of a Crossover Method using a Probabilistic Model for interactive Genetic Algorithm." In: *SCIS & ISIS SCIS & ISIS 2008*. Japan Society for Fuzzy Theory and Intelligent Informatics (cit. on p. 176).

Ho, Vivian (2019). *Bloody brilliant: new emoji to symbolize menstruation welcomed*. `https://www.theguardian.com/technology/2019/feb/09/period-emoji-menstruation-blood-donation`. Accessed: July 2021 (cit. on p. 146).

Hoffman, Guy and Gil Weinberg (2010). "Shimon: an interactive improvisational robotic marimba player." In: *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pp. 3097–3102 (cit. on p. 47).

Hong, Seunghoon, Dingdong Yang, Jongwook Choi, and Honglak Lee (2019). "Interpretable Text-to-Image Synthesis with Hierarchical Semantic Layout Generation." In: *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer, pp. 77–95 (cit. on p. 68).

Horton, William K (1994). *The icon book: Visual symbols for computer systems and documentation*. John Wiley & Sons, Inc. (cit. on pp. 1, 26, 27, 29, 30, 34–36, 90).

Hu, Tianran, Han Guo, Hao Sun, Thuy-vy Thi Nguyen, and Jiebo Luo (2017). "Spice up Your Chat: The Intentions and Sentiment Effects of Using Emoji." In: *arXiv preprint arXiv:1703.02860* (cit. on pp. 145, 149).

Huang, Forrest and John F Canny (2019). "Sketchforme: Composing sketched scenes from text descriptions for interactive applications." In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 209–220 (cit. on p. 66).

Huang, Forrest, Eldon Schoop, David Ha, and John Canny (2020). "Scones: towards conversational authoring of sketches." In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 313–323 (cit. on p. 66).

Huang, Hua, Lei Zhang, and Hong-Chao Zhang (2011). "Arcimboldo-like collage using internet images." In: *Proceedings of the 2011 SIGGRAPH Asia Conference*, pp. 1–8 (cit. on pp. 56, 57).

Hurtienne, Jörn (2009). "Image schemas and design for intuitive use: new guidance for user interface design." PhD thesis. Technische Universität Berlin (cit. on pp. 294–296).

Hyman, John (2006). *The objective eye*. University of Chicago Press (cit. on p. 87).

Iizuka, Yuichi, Hisako Shiohara, Tetsuya Iizuka, and Seiji Isobe (1998). "Automatic visualization method for visual data mining." In: *Pacific-Asia Conf. on Knowledge Discovery and Data Mining*. Springer (cit. on p. 242).

Inaba, Sho, Asako Kanezaki, and Tatsuya Harada (2014). "Automatic image synthesis from keywords using scene context." In: *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 1149–1152 (cit. on p. 56).

Indurkhya, Bipin and Amitash Ojha (2013). "An empirical study on the role of perceptual similarity in visual metaphors and creativity." In: *Metaphor and Symbol* 28.4, pp. 233–253 (cit. on p. 86).

– (2017). "Interpreting visual metaphors: asymmetry and reversibility." In: *Poetics Today* 38.1, pp. 93–121 (cit. on pp. 82, 86, 87).

Jahanian, Ali (2016a). "Design Mining Color Semantics." In: *Quantifying Aesthetics of Visual Design Applied to Automatic Design*. Springer. Chap. 3, pp. 15–55 (cit. on p. 2).

Jahanian, Ali (2016b). "Design Mining Visual Balance." In: *Quantifying Aesthetics of Visual Design Applied to Automatic Design*. Springer. Chap. 4, pp. 57–68 (cit. on p. 2).

Jain, Puneet, Najma Mathema, Jonathan Skaggs, and Dan Ventura (2021). "Ideation via Critic-Based Exploration of Generator Latent Space." In: *Proceedings of the 12th International Conference on Computational Creativity*. Ed. by Andrés Gómez de Silva Garza, Tony Veale, Wendy Aguilar, and Rafael Pérez y Pérez. México City, México (Virtual): Association for Computational Creativity, pp. 377–385. ISBN: 978-989-54160-3-5. URL: https://computationalcreativity.net/iccc21/wp-content/uploads/2021/09/ICCC_2021_paper_62.pdf (cit. on p. 64).

Jansen, Wim (2009). "Neurath, Arntz and ISOTYPE: the legacy in art, design and statistics." In: *Journal of Design History* 22.3, pp. 227–242 (cit. on p. 37).

Jespersen, Bjørn and Chris Reintges (2008). "Tractarian Sätze, Egyptian hieroglyphs, and the very idea of script as picture." In: *The philosophical forum*. Vol. 39. 1. Blackwell Publishing Inc Malden, USA, pp. 1–19 (cit. on p. 144).

Jiang, Yu, Jing Liu, and Hanqing Lu (2016). "Chat with illustration." In: *Multimedia Systems* 22.1, pp. 5–16 (cit. on p. 55).

Jobin, Anna, Marcello Ienca, and Effy Vayena (2019). "The global landscape of AI ethics guidelines." In: *Nature Machine Intelligence* 1.9, pp. 389–399 (cit. on p. 259).

Johnson, Mark (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press (cit. on pp. 287, 288).

Jongejan, J., H. Rowley, T. Kawashima, J. Kim, and N. Fox-Gieg (2016). *The Quick, Draw! - A.I. Experiment*. https://quickdraw.withgoogle.com/. Accessed: Oct. 2018 (cit. on pp. 65, 66).

Jordanous, Anna (2012). "A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative." In: *Cognitive Computation* 4.3, pp. 246–279 (cit. on p. 45).

– (2017). "Co-creativity and perceptions of computational agents in co-creativity." In: *Proceedings of the Eighth International Conference on Computational Creativity, Atlanta, US. ACC* (cit. on pp. 47, 205).

Joy, Annamma, John F. Sherry Jr., and Jonathan Deschenes (2009). "Conceptual blending in advertising." In: *Journal of Business Research* 62.1, pp. 39–49 (cit. on p. 275).

Kantosalo, Anna and Anna Jordanous (2020). "Role-Based Perceptions of Computer Participants in Human-Computer Co-Creativity." In: *7th Computational Creativity Symposium at AISB*. AISB (cit. on p. 48).

Karimi, Pegah, Kazjon Grace, Mary Lou Maher, and Nicholas Davis (2018a). "Evaluating Creativity in Computational Co-Creative Systems." In: *Proceedings of the Ninth International Conference on Computational Creativity* (cit. on pp. 46, 48, 205, 206).

Karimi, Pegah, Mary Lou Maher, Kazjon Grace, and Nicholas Davis (2018b). "A computational model for visual conceptual blends." In: *IBM Journal of Research and Development* (cit. on pp. 2, 66, 236, 274, 279).

Kasai, Hiroyuki (2013). "Lyric-based automatic music image generator for music browser using scene knowledge." In: *IEEE Transactions on Consumer Electronics* 59.3, pp. 578–586 (cit. on p. 55).

Kaye, Edward B (2001). "Good Flag, Bad Flag, and the Great NAVA Flag Survey of 2001." In: *Raven: A Journal of Vexillology* 8, pp. 11–38 (cit. on p. 256).

Keane, Mark T. and Fintan J. Costello (2001). "Setting limits on analogy: Why conceptual combination is not structural alignment." In: *The Analogical Mind: A Cognitive Science Perspective*. Ed. by D. Gentner, K.J. Holyoak, and B. Kokinov. Cambridge, MASS: MIT Press (cit. on p. 124).

Kelly, Ryan and Leon Watts (2015). "Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships." In: *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design* (cit. on p. 145).

Kennedy, John M (1982). "Metaphor in pictures." In: *Perception* 11.5, pp. 589–605 (cit. on pp. 22, 82).

Keogh, Eamonn, Li Wei, Xiaopeng Xi, Stefano Lonardi, Jin Shieh, and Scott Sirowy (2006). "Intelligent icons: Integrating lite-weight data mining and visualization into GUI operating systems." In: *Data Mining, 2006. ICDM'06. Sixth International Conference on*. IEEE, pp. 912–916 (cit. on p. 50).

Kikuchi, Go and Hiroyuki Kasai (2012). "Lyrics-based automatic music image generation using scene knowledge for music browsing." In: *2012 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, pp. 249–250 (cit. on p. 55).

Kim, Jingoog and Mary Lou Maher (2021). "Evaluating the Effect of Co-Creative Systems on Design Ideation." In: *Second Workshop on the Future of Co-Creative Systems at the 12th International Conference on Computational Creativity*. Ed. by Andrés Gómez de Silva Garza, Tony Veale, Wendy Aguilar, and Rafael Pérez y Pérez. México City, México (Virtual): Association for Computational Creativity, pp. 440–443 (cit. on p. 48).

Koestler, Arthur (1964). *The Act of Creation*. New York:Macmillan (cit. on p. 73).

Köhler, Wolfgang (1929). *Gestalt Psychology*. New York: Horace Liveright (cit. on pp. 14, 22).

Kowalewski, Hubert (2008). "Conceptual blending and sign formation." In: *The Public Journal of Semiotics* 2.2, pp. 30–51 (cit. on p. 282).

Kramer, Oliver (2010). "Evolutionary self-adaptation: a survey of operators and strategy parameters." In: *Evolutionary Intelligence* 3 (cit. on p. 206).

Krampen, Martin, Michael Götte, and Michael Kneidl (2007). *Die Welt der Zeichen: Kommunikation mit Piktogramme*. Ludwigsburg: avedition GmbH (cit. on p. 24).

Krcadinac, Uros, Jelena Jovanovic, Vladan Devedzic, and Philippe Pasquier (2015). "Textual affect communication and evocation using abstract generative visuals." In: *IEEE Transactions on Human-Machine Systems* 46.3, pp. 370–379 (cit. on p. 52).

Krzeczkowska, Anna, Jad El-Hage, Simon Colton, and Stephen Clark (2010). "Automated Collage Generation-With Intent." In: *ICCC*, pp. 36–40 (cit. on p. 56).

Kuhn, Werner (2007). "An image-schematic account of spatial categories." In: *International Conference on Spatial Information Theory*. Springer, pp. 152–168 (cit. on pp. 287–291).

Labov, William (1973). "The Boundaries of Words and Their Meanings." In: *New ways of analyzing variation in English*. Ed. by James N. Bailey and Roger W. Shuy. Washington DC: Georgetown Univeristy Press. (cit. on pp. 15, 17, 18).

Lakatos, I. (1976). *Proofs and refutations: the logic of mathematical discovery*. Cambridge University Press (cit. on p. 61).

Lakoff, George (1990). *Women, fire, and dangerous things*. University of Chicago press (cit. on pp. 15–18, 33).

Lamb, Carolyn (2018). "TwitSong: A current events computer poet and the thorny problem of assessment." PhD thesis. University of Waterloo (cit. on p. 45).

Lamy, Jean-Baptiste, Catherine Duclos, Avner Bar-Hen, Patrick Ouvrard, and Alain Venot (2008). "An iconic language for the graphical representation of medical concepts." In: *BMC medical informatics and decision making* 8.1, p. 16 (cit. on pp. 38, 83).

Lavater, Warja (1965). *Le Petit Chaperon Rouge*. A. Maeght (cit. on p. 29).

Lavin, Irving (1993). "Picasso's Bull(s): Art History in Reverse." In: *Art in America* (cit. on p. 28).

Leake, Mackenzie, Hijung Valentina Shin, Joy O Kim, and Maneesh Agrawala (2020). "Generating Audio-Visual Slideshows from Text Articles Using Word Concreteness." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–11 (cit. on p. 55).

Lebduska, Lisa (2014). *Emoji, emoji, what for art thou?* (Cit. on pp. 144, 150).

Leborg, Christian (2006). *Visual grammar*. Princeton Architectural Press (cit. on pp. 30, 32, 33).

Lee, Michael D, Rachel E Reilly, and Marcus E Butavicius (2003). "An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data." In: *Proc. of the Asia-Pacific symposium* *on Information visualisation-Volume 24*. Australian Computer Society, Inc. (cit. on p. 242).

Lee, Yong Jae, C Lawrence Zitnick, and Michael F Cohen (2011). "Shadowdraw: real-time user guidance for freehand drawing." In: *ACM Transactions on Graphics (TOG)*. Vol. 30. 4. ACM, p. 27 (cit. on pp. 46, 175).

Lelis, Catarina (2021). "Smart logos: a user's dashboard for the visualisation of meaningful brand experience data." In: *InfoDesign - Brazilian Journal of Information Design* 18.3, pp. 85–104 (cit. on p. 41).

Lewis, John P, Ruth Rosenholtz, Nickson Fong, and Ulrich Neumann (2004). "VisualIDs: automatic distinctive icons for desktop interfaces." In: *ACM Transactions on Graphics (TOG)*. Vol. 23. 3. ACM, pp. 416–423 (cit. on p. 50).

Li, Tzu-Mao, Michal Lukáč, Michaël Gharbi, and Jonathan Ragan-Kelley (2020). "Differentiable vector graphics rasterization for editing and learning." In: *ACM Transactions on Graphics (TOG)* 39.6, pp. 1–15 (cit. on p. 65).

Li, X. Alice and Devi Parikh (2020). "Lemotif: An Affective Visual Journal Using Deep Neural Networks." In: *Proceedings of the Eleventh International Conference on Computational Creativity* (cit. on p. 62).

Li, Yi-Na, Dong-Jin Li, and Kang Zhang (2015). "Metaphoric transfer effect in information visualization using glyphs." In: *Proc. of the 8th Int. Symposium on Visual Information Communication and Interaction*. ACM, pp. 121–130 (cit. on pp. 241, 248, 249).

Liapis, Antonios (2018). "Recomposing the Pokémon Color Palette." In: *Applications of Evolutionary Computation*. Springer (cit. on p. 61).

Liapis, Antonios, Amy K. Hoover, Georgios N. Yannakakis, Constantine Alexopoulos, and Evangelia V. Dimaraki (2015). "Motivating visual interpretations in iconoscope: Designing a game for fostering creativity." In: *Proceedings of the Foundations of Digital Games Conference* (cit. on p. 53).

Liapis, Antonios, Georgios N Yannakakis, Constantine Alexopoulos, and Phil Lopes (2016). "Can Computers Foster Human Users' Creativity? Theory and Praxis of Mixed-Initiative Co-Creativity." In: *Digital Culture & Education* 8.2, pp. 136–153 (cit. on pp. 3, 48, 236).

Liapis, Antonios, Georgios N Yannakakis, and Julian Togelius (2013). "Sentient Sketchbook: Computer-aided game level authoring." In: *FDG*, pp. 213–220 (cit. on p. 180).

Lin, Sharon et al. (2013). "Selecting Semantically-Resonant Colors for Data Visualization." In: *Computer Graphics Forum*. Vol. 32. 3pt4. Wiley Online Library, pp. 401–410 (cit. on pp. 2, 23, 52).

Lin, Ting-Ju and Michael Biggs (2006). "A preliminary study of learnable pictogram languages." In: *Design Research Society International Conference Proceedings*. IADE (cit. on pp. 37, 38).

Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick (2014).

"Microsoft coco: Common objects in context." In: *European conference on computer vision*. Springer, pp. 740–755 (cit. on p. 300).
Liu, Hugo and Push Singh (2004). "ConceptNet—a practical commonsense reasoning tool-kit." In: *BT technology journal* 22.4, pp. 211–226 (cit. on p. 90).
Liu, Yiming, Aseem Agarwala, Jingwan Lu, and Szymon Rusinkiewicz (2016). "Data-driven iconification." In: *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*. Eurographics Association, pp. 113–124 (cit. on p. 59).
Llano, Maria Teresa, Mark d'Inverno, Matthew Yee-King, Jon McCormack, Alon Ilsar, Alison Pease, and Simon Colton (2020). "Explainable Computational Creativity." In: *Proceedings of the Eleventh International Conference on Computational Creativity* (cit. on pp. 260, 327).
Long, Duri (2021). "Designing Co-Creative, Embodied AI Literacy Interventions for Informal Learning Spaces." PhD thesis. Georgia Institute of Technology (cit. on p. 45).
Long, Duri, Sanjana Gupta, Jessica Brooke Anderson, and Brian Magerko (2017a). "The Shape of Story: A Semiotic Artistic Visualization of a Communal Storytelling Experience." In: *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference* (cit. on p. 53).
Long, Duri, Mikhail Jacob, Nicholas Davis, and Brian Magerko (2017b). "Designing for socially interactive systems." In: *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. ACM, pp. 39–50 (cit. on p. 47).
Lopes, Carolina Gonçalves, João Miguel Cunha, and Pedro Martins (2020). "Towards Generative Illustration of Text." In: *Joint Proceedings of the ICCC 2020 Workshops (WS 2020)*. Ed. by Max Kreminski, Viktor Eisenstadt, Sofia Pinto, and Oliver Kutz (cit. on pp. 6, 91, 329).
Loughran, Róisín and Michael O'Neill (2017). "Application Domains Considered in Computational Creativity." In: *Proceedings of the Eighth International Conference on Computational Creativity*, pp. 197–204 (cit. on p. 45).
Lourenço, Nuno, Filipe Assunção, Catarina Maçãs, and Penousal Machado (2017). "EvoFashion: Customising Fashion Through Evolution." In: *Int. Conf. on Evolutionary and Biologically Inspired Music and Art*. Springer International Publishing (cit. on pp. 176, 180, 181, 186).
Lubart, Todd (Oct. 2005). "How can computers be partners in the creative process: Classification and commentary on the Special Issue." In: *International Journal of Human-Computer Studies* 63.4-5, pp. 365–369. DOI: 10.1016/j.ijhcs.2005.04.002. URL: https://doi.org/10.1016/j.ijhcs.2005.04.002 (cit. on p. 47).
Lucas, Gavin (2016). *The story of emoji*. Prestel Verlag (cit. on p. 143).
Lyman, Bernard (1979). "Representation of complex emotional and abstract meanings by simple forms." In: *Perceptual and Motor Skills* 49.3, pp. 839–842 (cit. on p. 22).

MacEachren, Alan M (2001). "An evolving cognitive-semiotic approach to geographic visualization and knowledge construction." In: *Information Design Journal* 10.1, pp. 26–36 (cit. on p. 13).

MacGregor, Donald and Paul Slovic (1986). "Graphic representation of judgmental information." In: *Human-Computer Interaction* 2.3 (cit. on p. 249).

MacLeod, Colin M (1991). "Half a century of research on the Stroop effect: an integrative review." In: *Psychological bulletin* 109.2, p. 163 (cit. on p. 23).

Maçãs, Catarina, Nuno Lourenço, and Penousal Machado (2018). "Interactive Evolution of Swarms for the Visualisation of Consumptions." In: *ArtsIT 2018* (cit. on p. 176).

Maçãs, Catarina, David Palma, and Artur Rebelo (2019). "typEm: a generative typeface that represents the emotion of the text." In: *Proceedings of the 9th International Conference on Digital and Interactive Arts*, pp. 1–10 (cit. on pp. 51, 52).

Machado, Daniel Leal Moreira (2018). "Construção de modelos neurais para criação de arte generativa visual." PhD thesis. Master thesis, University of Porto (cit. on p. 64).

Machado, Fernando Jorge Penousal Martins (2007). "Inteligencia artificial e arte." PhD thesis. University of Coimbra (cit. on p. 17).

Machado, Penousal, João Correia, and Juan Romero (2012). "Expression-based evolution of faces." In: *International Conference on Evolutionary and Biologically Inspired Music and Art*. Springer, pp. 187–198 (cit. on p. 49).

Machado, Penousal, Adriano Vinhas, João Correia, and Aniko Ekárt (2015). "Evolving ambiguous images." In: *AI Matters* 2.1, pp. 7–8 (cit. on pp. 50, 80).

Machida, Wakako and Takayuki Itoh (2011). "Lyricon: A visual music selection interface featuring multiple icons." In: *15th International Conference on Information Visualisation*. IEEE, pp. 145–150 (cit. on p. 55).

Mackinlay, Jock, Pat Hanrahan, and Chris Stolte (2007). "Show me: Automatic presentation for visual analysis." In: *IEEE Trans. on visualization and computer graphics* 13.6 (cit. on p. 243).

Maes, Alfons and Joost Schilperoord (2008). "Conceptual and Structural Heuristics." In: *Go Figure! New Directions in Advertising Rhetoric* 227 (cit. on p. 85).

Maher, Mary Lou (2012). "Computational and Collective Creativity: Who's Being Creative?" In: *International Conference on Computational Creativity*, p. 67 (cit. on p. 47).

Mamykina, Lena, Linda Candy, and Ernest Edmonds (2002). "Collaborative creativity." In: *Communications of the ACM* 45.10, pp. 96–99 (cit. on p. 47).

Mandler, Jean M (2000). "Perceptual and conceptual processes in infancy." In: *Journal of cognition and development* 1.1, pp. 3–36 (cit. on pp. 18, 287).

Mano, Tetsuaki, Hiroaki Yamane, and Tatsuya Harada (2016). “Scene Image Synthesis from Natural Sentences Using Hierarchical Syntactic Analysis.” In: *Proceedings of the 24th ACM international conference on Multimedia*, pp. 112–116 (cit. on p. 56).

Mansimov, Elman, Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov (2015). “Generating images from captions with attention.” In: *arXiv preprint arXiv:1511.02793* (cit. on p. 67).

Marinthe, Gaëlle, Juan Manuel Falomir-Pichastor, Benoit Testé, and Rodolphe Kamiejski (2019). “Flags on fire: Consequences of a national symbol’s desecration for intergroup relations.” In: *Group Processes & Intergroup Relations* (cit. on p. 267).

Marr, David (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman and Co. (cit. on p. 12).

Martins, P, S Pollak, T Urbancic, and A Cardoso (2016). “Optimality Principles in Computational Approaches to Conceptual Blending: Do We Need Them (at) All?” In: *Proceedings of the Seventh International Conference on Computational Creativity* (cit. on pp. 139, 283).

Martins, P., T. Urbancic, S. Pollak, N. Lavrac, and A Cardoso (2015). “The Good, the Bad, and the AHA! Blends.” In: *6th International Conference on Computational Creativity, ICCC 2015* (cit. on pp. 87, 108, 275, 276, 281–283).

Martins, Pedro, H Gonçalo Oliveira, João Carlos Gonçalves, António Cruz, F Amílcar Cardoso, Martin Žnidaršič, Nada Lavrač, Simo Linkola, Hannu Toivonen, Raquel Hervás, et al. (2019a). “Computational creativity infrastructure for online software composition: A conceptual blending use case.” In: *IBM Journal of Research and Development* 63.1, pp. 9–1 (cit. on p. 284).

Martins, Tiago (2021). “Automated Evolution for Design.” PhD thesis. University of Coimbra (cit. on pp. 43, 44).

Martins, Tiago, João Correia, Ernesto Costa, and Penousal Machado (2019b). “Evolving Stencils for Typefaces: Combining Machine Learning, User’s Preferences and Novelty.” In: *Complexity* 2019, 3509263:1–3509263:16. DOI: 10.1155/2019/3509263. URL: https://doi.org/10.1155/2019/3509263 (cit. on pp. 41, 43).

Martins, Tiago, João Miguel Cunha, João Bicker, and Penousal Machado (2019c). “Dynamic Visual Identities: from a survey of the state-of-the-art to a model of features and mechanisms.” In: *Visible Language* 53.2, pp. 4–35 (cit. on pp. 43, 257, 258).

Matusitz, J (2007). “Vexillology, or how flags speak.” In: *International Journal of Applied Semiotics* 5.1, pp. 199–211 (cit. on p. 264).

McCaig, Graeme, Steve DiPaola, and Liane Gabora (2016). “Deep Convolutional Networks as Models of Generalization and Blending Within Visual Creativity.” In: *Proceedings of the Seventh International Conference on Computational Creativity* (cit. on p. 64).

McCloud, Scott (1993). *Understanding comics: The invisible art*. Kitchen Sink Press (cit. on pp. 28, 35).

McCorduck, Pamela (1991). *Aaron's code: meta-art, artificial intelligence, and the work of Harold Cohen*. Macmillan (cit. on p. 175).

McLuhan, Marshall (1962). *The Gutenberg Galaxy: The Making of the Typographic Man*. Toronto: University of Toronto Press (cit. on p. 268).

McQuarrie, Edward F (2008). "Differentiating the pictorial element in advertising: A rhetorical perspective." In: *Visual marketing: From attention to action.* Lawrence Erlbaum Associates, Taylor & Francis Group (cit. on pp. 1, 81, 84, 88).

Medin, Douglas L and Marguerite M Schaffer (1978). "Context theory of classification learning." In: *Psychological review* 85.3, p. 207 (cit. on p. 15).

Meggs, Philip B and Alston W Purvis (2012). *Meggs' History of Graphic Design*. Fifth edition. John Wiley & Sons, Inc., Hoboken, New Jersey (cit. on p. 1).

Mei, Tao, Wei Zhang, and Ting Yao (2020). "Vision and language: from visual perception to content creation." In: *APSIPA Transactions on Signal and Information Processing* 9 (cit. on p. 67).

Mellor, John FJ, Eunbyung Park, Yaroslav Ganin, Igor Babuschkin, Tejas Kulkarni, Dan Rosenbaum, Andy Ballard, Theophane Weber, Oriol Vinyals, and SM Eslami (2019). "Unsupervised doodling and painting with improved spiral." In: *arXiv preprint arXiv:1910.01007* (cit. on p. 64).

Mendel, Gregor (1865). *Experiments in Plant Hybridization*. Oliver & Boyd. (cit. on p. 175).

Mihai, Daniela and Jonathon Hare (2021). "Learning to Draw: Emergent Communication through Sketching." In: *arXiv preprint arXiv:2106.02067* (cit. on p. 66).

Miller, Hannah, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht (2016). "Blissfully happy"or "ready to fight": Varying Interpretations of Emoji." In: *Proc. of ICWSM* 2016 (cit. on pp. 145, 155, 234, 305).

Minsky, Marvin (1985). *The society of mind*. New York: Simon and Schuster (cit. on p. 30).

Mittal, Paritosh, Kunal Aggarwal, Pragya Paramita Sahu, Vishal Vatsalya, Soumyajit Mitra, Vikrant Singh, Viswanath Veera, and Shankar M Venkatesan (2020). "Photo-realistic emoticon generation using multi-modal input." In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 254–258 (cit. on pp. 68, 145, 148).

Morales-Ramirez, Carlos A (2018). "Measuring Puerto Ricans' knowledge of the national, subnational and Latin American flags." In: *Research in Social Sciences and Technology* 3.3, pp. 42–67 (cit. on pp. 264, 266).

Mordvintsev, Alexander, Christopher Olah, and Mike Tyka (2015). *Inceptionism: Going Deeper into Neural Networks*. URL: `https://research.`

googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html (cit. on p. 63).

Murphy, Gregory L and Douglas L Medin (1985). “The role of theories in conceptual coherence.” In: *Psychological review* 92.3, p. 289 (cit. on p. 15).

Nakano, Reiichiro (2019). “Neural painters: A learned differentiable constraint for generating brushstroke paintings.” In: *arXiv preprint arXiv:1904.08410* (cit. on p. 65).

Napier, Mark (2002). “NET.FLAG.” In: *Ars Electronica 2002: UNPLUGGED – Art as the Scene of Global Conflicts*. Ed. by Gerfried Stocker and Christine Schöpf. Hatje Cantz Publishers, pp. 360–361 (cit. on pp. 252, 253, 255, 257).

Navigli, Roberto and Simone Paolo Ponzetto (2012). “BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network.” In: *Artificial Intelligence* 193, pp. 217–250 (cit. on p. 161).

Neff, Jack (2015). *Dove Launches Curly-haired Emojis To End Straight-hair Dominance: Brand Rescues Wavy Haired 'marginalized' By Emojis*. https://adage.com/article/digital/dove-launches-curly-haired-emojis-address-void/301203. Accessed: July 2021 (cit. on p. 146).

Negro, Isabel, Ester Šorm, and Gerard Steen (2018). “General image understanding in visual metaphor identification.” In: *Odisea nº 18: Revista de estudios ingleses* 18, p. 113 (cit. on p. 1).

Nelson, Elisabeth, David Dow, Christopher Lukinbeal, and Ray Farley (1997). “Visual search processes and the multivariate point symbol.” In: *Cartographica: The Int. Journal for Geographic Information and Geovisualization* 34.4, pp. 19–33 (cit. on pp. 247, 249).

Neurath, Otto (1936). *International Picture Language. The First Rules of Isotype... With Isotype Pictures*. Kegan Paul & Company (cit. on p. 37).

Nguyen, Anh, Alexey Dosovitskiy, Jason Yosinski, Thomas Brox, and Jeff Clune (2016). “Synthesizing the preferred inputs for neurons in neural networks via deep generator networks.” In: *Advances in neural information processing systems* 29, pp. 3387–3395 (cit. on p. 63).

Niediek, Imke (2016). “Don't write it, picture it!: Accessible Information by graphic signs.” In: *Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*. ACM, pp. 188–193 (cit. on p. 149).

Norton, David, Derrall Heath, and Dan Ventura (2011). “Autonomously Creating Quality Images.” In: *ICCC*. Citeseer, pp. 10–15 (cit. on pp. 62, 63).

– (2013). “Finding creativity in an artificial artist.” In: *The Journal of Creative Behavior* 47.2, pp. 106–124 (cit. on p. 47).

Novak, Petra Kralj, Jasmina Smailović, Borut Sluban, and Igor Mozetič (2015). “Sentiment of emojis.” In: *PloS one* 10.12 (cit. on p. 145).

Ojha, Amitash, Charles Forceville, and Bipin Indurkhya (2021). "An experimental study on the effect of emotion lines in comics." In: *Semiotica* (cit. on p. 35).

Ojha, Amitash and Bipin Indurkhya (2020). "On the role of perceptual similarity in producing visual metaphors." In: *Producing Figurative Expression: Theoretical, experimental and practical perspectives* 10, p. 105 (cit. on p. 82).

Oliva, Aude (2013). "The art of hybrid images: Two for the view of one." In: *Art & Perception* 1.1-2, pp. 65–74 (cit. on pp. 79, 80).

Onsager, Alex (2013). *Pokemon Fusion: Behind the Scenes*. https://www.alexonsager.com/2013/06/04/behind-the-scenes-pokemon-fusion.html. [Online; accessed Dec. 2021] (cit. on pp. 61, 62).

Ortiz, María J (2010). "Visual rhetoric: Primary metaphors and symmetric object alignment." In: *Metaphor and Symbol* 25.3, pp. 162–180 (cit. on p. 86).

Osborne, Danny, Jennifer Lees-Marshment, Clifton van der Linden, and Others (2016). "National identity and the flag change referendum: Examining the latent profiles underlying New Zealanders' flag change support." In: *New Zealand Sociology* 31.7, p. 19 (cit. on p. 267).

Padró, Lluís and Evgeny Stanilovsky (2012). "Freeling 3.0: Towards wider multilinguality." In: *LREC2012* (cit. on p. 194).

Parmee, Ian C, Johnson AR Abraham, and Azahar Machwe (2008). "User-centric evolutionary computing: Melding human and machine capability to satisfy multiple criteria." In: *Multiobjective Problem Solving from Nature*. Springer, pp. 263–283 (cit. on p. 176).

Patashnik, Or, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski (2021). "Styleclip: Text-driven manipulation of stylegan imagery." In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2085–2094 (cit. on p. 68).

Pater, Ruben (2012). *Double Standards*. Acter (cit. on pp. 252, 253, 266, 267, 269).

Pelikan, Martin, David E Goldberg, and Fernando G Lobo (2002). "A survey of optimization by building and using probabilistic models." In: *Computational optimization and applications* (cit. on p. 176).

Pereira, Francisco C. (2007). *Creativity and Artificial Intelligence: A Conceptual Blending Approach*. Berlin: Mouton de Gruyter (cit. on p. 60).

Pereira, Francisco C and Amílcar Cardoso (2002). "The boat-house visual blending experience." In: *Proceedings of the Symposium for Creativity in Arts and Science of AISB 2002* (cit. on pp. 58–60, 121, 130, 274, 279, 322).

Pereira, Francisco Câmara (2004). "Um Modelo Computacional de Criatividade." PhD thesis. University of Coimbra (cit. on pp. 14, 78, 123, 124).

Pérez, Patrick, Michel Gangnet, and Andrew Blake (2003). "Poisson image editing." In: *ACM SIGGRAPH 2003 Papers*, pp. 313–318 (cit. on p. 81).

Pérez, Rafael Pérez y (2018). "The Computational Creativity Continuum." In: *Proceedings of the Ninth International Conference on Computational Creativity, ICCC 2018, Salamanca, Spain, June 25-29, 2018*. Ed. by François Pachet, Anna Jordanous, and Carlos León. Association for Computational Creativity (ACC), pp. 177–184 (cit. on p. 46).

Petajan, Eric D, Yves D Jean, Dan Lieuwen, and Vinod Anupam (1997). "Dataspace: An automated visualization system for large databases." In: *Visual Data Exploration and Analysis IV*. Vol. 3017. Int. Society for Optics and Photonics, pp. 89–99 (cit. on p. 242).

Peterson, H Philip (1965). "The Digital Mona Lisa." In: *Computers and Automation* 14 (cit. on p. 49).

Peterson, Matthew O (2018). "Aspects of visual metaphor: an operational typology of visual rhetoric for research in advertising." In: *International Journal of Advertising* 38.1, pp. 67–96 (cit. on pp. 81, 82, 84, 85, 88, 94–97, 107, 114, 115, 276).

Petridis, Savvas and Lydia B. Chilton (2019). "Human Errors in Interpreting Visual Metaphor." In: *Proceedings of the 2019 ACM SIGCHI Conference on Creativity and Cognition, C&C 2019, San Diego, CA, USA, June 23-26, 2019*. Pp. 187–197. URL: https://doi.org/10.1145/3325480.3325503 (cit. on pp. 82, 275).

Pettersson, Rune (2011). *Information Design, Volume 5: Cognition*. Institute for infology (cit. on pp. 12, 14).

Phillips, Barbara J and Edward F McQuarrie (2004). "Beyond visual metaphor: A new typology of visual rhetoric in advertising." In: *Marketing theory* 4.1-2, pp. 113–136 (cit. on pp. 61, 83–85, 88, 89, 163).

Piper, Adam Kelly (2010). "Participatory design of warning symbols using distributed interactive evolutionary computation." PhD thesis. Auburn University (cit. on pp. 53, 176).

Plutchik, Robert (2001). "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice." In: *American scientist* 89.4, pp. 344–350 (cit. on p. 52).

Pohl, Henning, Christian Domin, and Michael Rohs (2017). "Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication." In: *ACM TOCHI-17* 24.1, p. 6 (cit. on p. 145).

Polisciuc, Evgheni (2021). "Thematic Cartography for Adaptive Visualization Systems." PhD thesis. University of Coimbra (cit. on p. 13).

Pollak, S., P. Martins, A. Cardoso, and T. Urbancic (2015). "Automated blend naming based on human creativity examples." In: *23rd International Conference on Case-Based Reasoning (ICCBR 2015) – Workshop on Experience and Creativity* (cit. on pp. 275, 284, 324).

Poppi, FIM, M Bolognesi, and A Ojha (2020). "Imago Dei: Metaphorical conceptualization of pictorial artworks within a participant-based framework." In: *Semiotica* (cit. on p. 98).

Prada, Marília, David Rodrigues, Rita R Silva, and Margarida V Garrido (2016). "Lisbon symbol database (LSD): subjective norms for

600 symbols." In: *Behavior research methods* 48.4, pp. 1370–1382 (cit. on pp. 196, 304).

Puyat, Marcel (2017). "EmotiGAN: Emoji Art using Generative Adversarial Networks." CS229: Machine Learning Course, Stanford University (cit. on pp. 68, 145).

Qiu, Xigui (2000). "Chinese Writing. Society for the Study of Early China and Institute of East Asian Studies." In: *University of California, Berkeley* (cit. on p. 36).

Quendler, Christian (2016). *The Camera-eye metaphor in cinema*. Routledge (cit. on p. 20).

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. (2021). "Learning transferable visual models from natural language supervision." In: *arXiv preprint arXiv:2103.00020* (cit. on pp. 65, 68, 69).

Radpour, Dianna and Vivek Bheda (2017). "Conditional Generative Adversarial Networks for Emoji Synthesis with Word Embedding Manipulation." In: *arXiv preprint arXiv:1712.04421* (cit. on pp. 64, 145).

Rafner, Janet, Arthur Hjorth, Sebastian Risi, Lotte Philipsen, Charles Dumas, Michael Mose Biskjær, Lior Noy, Kristian Tylén, Carsten Bergenholtz, Jesse Lynch, et al. (2020). "crea. blender: A Neural Network-Based Image Generation Game to Assess Creativity." In: *CHI Play* (cit. on p. 67).

Raiola, Gaetano, Domenico Tafuri, and F Gomez Paloma (2014). "Physical activity and sport skills and its relation to mind theory on motor control." In: *Sport Science* 7.1, pp. 52–56 (cit. on p. 18).

Ramachandran, Vilayanur S and Edward M Hubbard (2003). "Hearing colors, tasting shapes." In: *Scientific American* 288.5, pp. 52–59 (cit. on p. 22).

Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever (2021). "Zero-shot text-to-image generation." In: *arXiv preprint arXiv:2102.12092* (cit. on p. 68).

Ravikumar, Prashanth Thattai (2020). "Looking Beyond Creative Support: Augmenting the Perceived Agency of Music Co-Creation Systems." PhD thesis. National University of Singapore. URL: `https://scholarbank.nus.edu.sg/handle/10635/167565` (cit. on p. 47).

Reale, Cesco, Marwan Kilani, Araceli Giménez, Nadu Barbashova, and Roman Oechslin (2021). "From Hieroglyphs to Emoji, to IKON: The Search of the (Perfect?) Visual Language." In: *Design, User Experience, and Usability: UX Research and Design - 10th International Conference, DUXU 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24-29, 2021, Proceedings, Part I*. Ed. by Marcelo M. Soares, Elizabeth Rosenzweig, and Aaron Marcus. Vol. 12779. Lecture Notes in Computer Science. Springer, pp. 457–476. DOI: `10.1007/978-3-030-78221-4\_31` (cit. on pp. 34, 36–38).

Reas, Casey and Chandler McWilliams (2010). *Form+ Code: in design, art, and architecture*. Princeton Architecturel Press (cit. on pp. 42, 43).

Rebelo, Sérgio and Carlos M Fonseca (2018). "Experiments in the Development of Typographical Posters." In: *6th Conf. on Computation, Communication, Aesthetics and X*) (cit. on p. 176).

Rebelo, Sérgio, Tiago Martins, João Bicker, and Penousal Machado (2018). "Typography as image: experiments on typographic portraits." In: *Proceedings of the 9th Typography Meeting* (cit. on p. 80).

Rebelo, Sérgio, Catarina Pires, Pedro Martins, João Bicker, and Machado Penousal (2019). "Designing Posters Towards a Seamless Integration in Urban Surroundings: A Computational Approach." In: *Proceedings of the 9th international Conference on Digital and Interactive Arts*. ACM (cit. on p. 258).

Reed, Scott E., Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee (2016). "Generative Adversarial Text to Image Synthesis." In: *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pp. 1060–1069. URL: http://proceedings.mlr.press/v48/reed16.html (cit. on pp. 64, 68).

Regier, Terry (1996). *The human semantic potential: Spatial language and constrained connectionism*. MIT Press (cit. on p. 33).

Rezwana, Jeba and Mary Lou Maher (2021). "COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems." In: *Proceedings of the 2nd Workshop on the Future of Co-Creative Systems* (cit. on p. 48).

Ribeiro, Paulo, Francisco C. Pereira, Bruno Marques, Bruno Leitao, and Amílcar Cardoso (2003). "A Model for Creativity in Creature Generation." In: *4th International Conference on Intelligent Games and Simulation (GAME-ON 2003)* (cit. on p. 61).

Richards, Clive James (1984). "Diagrammatics: an investigation aimed at providing a theoretical framework for studying diagrams and for establishing a taxonomy of their fundamental modes of graphic organization." PhD thesis. Royal College of Art (cit. on p. 26).

Richardson, Andrew (2017). *Data-driven Graphic Design: Creative Coding for Visual Communication*. Bloomsbury Publishing (cit. on pp. 41, 257).

Roberts, Will and Markus Egg (2018). "A large automatically-acquired all-words list of multiword expressions scored for compositionality." In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (cit. on pp. 76, 199, 226).

Robertson, Alexander, Farhana Ferdousi Liza, Dong Nguyen, Barbara McGillivray, and Scott A Hale (2021). "Semantic Journeys: Quantifying Change in Emoji Meaning from 2012-2018." In: *Proceedings of the 4th International Workshop on Emoji Understanding and Applications in Social Media (Emoji2021)* (cit. on p. 147).

Rodrigues, Ana, Amílcar Cardoso, and Penousal Machado (2019). "A Dynamic Approach for the Generation of Perceptual Associations." In: *Proceedings of the Tenth International Conference on Computational Creativity, UNC Charlotte, North Carolina, June 17-21, 2019.* (Cit. on p. 51).

Rodrigues, David, Marília Prada, Rui Gaspar, Margarida V Garrido, and Diniz Lopes (2018). "Lisbon emoji and emoticon database (LEED): norms for emoji and emoticons in seven evaluative dimensions." In: *Behavior research methods* 50.1, pp. 392–405 (cit. on pp. 145, 304).

Rosch, Eleanor (1975). "Cognitive representations of semantic categories." In: *Journal of experimental psychology: General* 104.3, p. 192. DOI: 10.1037/0096-3445.104.3.192 (cit. on p. 15).

Rosch, Eleanor and Carolyn B Mervis (1975). "Family resemblances: Studies in the internal structure of categories." In: *Cognitive psychology* 7.4, pp. 573–605 (cit. on pp. 15, 16).

Rosch, Eleanor, Carolyn B Mervis, Wayne D Gray, David M Johnson, and Penny Boyes-Braem (1976). "Basic objects in natural categories." In: *Cognitive psychology* 8.3, pp. 382–439 (cit. on pp. 15–17).

Roy, Deb K (2002). "Learning visually grounded words and syntax for a scene description task." In: *Computer speech & language* 16.3-4, pp. 353–385 (cit. on p. 33).

Roy, Rishiraj Saha, Abhijeet Singh, Prashant Chawla, Shubham Saxena, and Atanu R Sinha (2017). "Automatic assignment of topical icons to documents for faster file navigation." In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 1. IEEE, pp. 1338–1345 (cit. on p. 54).

Saint-Exupéry, Antoine de (1943). *The Little Prince*. Reynal & Hitchcock (cit. on p. 29).

Santos, André Tomás, Domingos J Cruz, and António Fernando Barbosa (2017). "Gravuras e pinturas em dólmenes. O 'grupo de Viseu' de E. Shee (1981) trinta anos depois." In: *Actas da Mesa-Redonda A Pré–história ea Proto-história no Centro de Portugal: avaliação e perspectivas de futuro*, pp. 25–57 (cit. on pp. 1, 11).

Sarajčić, Ivan (2007). "Flag Identifier - Flag Identifying Tool and Vexillological Database – An Attempt of Vexillological Classification." In: *22nd International Congress of Vexillology - FlagBerlin2007* (cit. on p. 255).

Satherley, Nicole, Danny Osborne, and Chris G Sibley (2019). "Who Is for (or Against) the National Flag? Ideological and Identity-Based Motivators of Attitudes." In: *Analyses of Social Issues and Public Policy* 19.1, pp. 407–428 (cit. on p. 265).

Sbai, Othman (2021). "Deep learning methods for visual content creation and understanding." PhD thesis. Marne-la-vallée, ENPC (cit. on pp. 60, 64).

Sbai, Othman, Camille Couprie, and Mathieu Aubry (2021). "Surprising Image Compositions." In: *Proceedings of the 12th International Con-*

ference on Computational Creativity. Ed. by Andrés Gómez de Silva Garza, Tony Veale, Wendy Aguilar, and Rafael Pérez y Pérez. México City, México (Virtual): Association for Computational Creativity, pp. 248–255. ISBN: 978-989-54160-3-5 (cit. on p. 60).

Schaldenbrand, Peter, Zhixuan Liu, and Jean Oh (2021). "StyleCLIP-Draw: Coupling Content and Style in Text-to-Drawing Synthesis." In: *arXiv preprint arXiv:2111.03133* (cit. on p. 68).

Schilperoord, Joost (2018). "Ways with pictures." In: *Visual Metaphor: Structure and process*. John Benjamins Publishing Company, pp. 11–46 (cit. on pp. 18, 82, 83, 94–96).

Schilperoord, Joost, Alfons Maes, and Heleen Ferdinandusse (2009). "Perceptual and conceptual visual rhetoric: The case of symmetric object alignment." In: *Metaphor and Symbol* 24.3, pp. 155–173 (cit. on pp. 86, 88).

Schwarz, Katharina, Tamara L Berg, and Hendrik PA Lensch (2016). "Auto-illustrating poems and songs with style." In: *Asian Conference on Computer Vision*. Springer, pp. 87–103 (cit. on p. 55).

Seiça, Mariana, Sérgio Rebelo, Nuno Lourenço, and Pedro Martins (2021). "#ESSYS*: An Online Happening." In: *xCoAx 2021: 9th Conference on Computation, Communication, Aesthetics & X*. Ed. by Luísa Ribas Miguel Carvalhais Mario Verdicchio and André Rangel. Porto, Portugal: Research Institute in Art, Design and Society School of Fine Arts (i2ADS), University of Porto, pp. 570–577 (cit. on p. 52).

Setlur, Vidya, Conrad Albrecht-Buehler, Amy A. Gooch, Sam Rossoff, and Bruce Gooch (2005). "Semanticons: Visual metaphors as file icons." In: *Computer Graphics Forum*. Vol. 24. 3. Wiley Online Library, pp. 647–656 (cit. on p. 54).

Setlur, Vidya and Jock D Mackinlay (2014). "Automatic generation of semantic icon encodings for visualizations." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 541–550 (cit. on p. 54).

Siangliulue, Pao, Kenneth C Arnold, Krzysztof Z Gajos, and Steven P Dow (2015). "Toward collaborative ideation at scale: Leveraging ideas from others to generate more creative and diverse ideas." In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 937–945 (cit. on p. 2).

Siirtola, Harri (2005). "The effect of data-relatedness in interactive glyphs." In: *Proc. of the Ninth Int. Conf. on Information Visualisation*. IEEE, pp. 869–876 (cit. on pp. 242, 246, 247).

Silvers, Robert (1996). "Photomosaics: putting pictures in their place." MA thesis. Massachusetts Institute of Technology (cit. on p. 80).

Sims, Karl (1991). "Artificial evolution for computer graphics." In: *Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, pp. 319–328 (cit. on p. 49).

Sittenfeld, Curtis and Jennifer Daniel (2014). *The Emojis We Really Need*. https://www.nytimes.com/interactive/2014/07/06/sunday-

review/The-Emojis-We-Really-Need.html. Accessed: July 2021 (cit. on p. 146).
Speer, Robert and Catherine Havasi (2012). "Representing General Relational Knowledge in ConceptNet 5." In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*, pp. 3679–3686 (cit. on pp. 90, 159, 330).
Stebbing, Peter D (2004). "A universal grammar for visual composition?" In: *Leonardo* 37.1, pp. 63–70 (cit. on p. 14).
Steinbrück, Alexa (2013). "Conceptual blending for the visual domain." PhD thesis. Master's thesis, University of Amsterdam (cit. on pp. 58, 61).
Stiny, George and James Gips (1971). "Shape grammars and the generative specification of painting and sculpture." In: *IFIP congress (2)*. Vol. 2. 3, pp. 125–135 (cit. on p. 43).
Sun, Lingyun, Pei Chen, Wei Xiang, Peng Chen, Wei-yue Gao, and Kejun Zhang (2019). "SmartPaint: a co-creative drawing system based on generative adversarial networks." In: *Frontiers of Information Technology & Electronic Engineering* 20.12, pp. 1644–1656 (cit. on p. 64).
Taigman, Yaniv, Adam Polyak, and Lior Wolf (2016). "Unsupervised cross-domain image generation." In: *arXiv preprint arXiv:1611.02200* (cit. on p. 148).
Takagi, Hideyuki (1998). "Interactive evolutionary computation: System optimization based on human subjective evaluation." In: *IEEE International Conference on Intelligent Engineering Systems*. Vol. 1998, pp. 17–19 (cit. on p. 44).
Tamés, David (2009). "Collaborative Visual Lexicon." In: (cit. on pp. 1, 323).
Tao, Ming, Hao Tang, Songsong Wu, Nicu Sebe, Xiao-Yuan Jing, Fei Wu, and Bingkun Bao (2020). "Df-gan: Deep fusion generative adversarial networks for text-to-image synthesis." In: *arXiv preprint arXiv:2008.05865* (cit. on p. 70).
Tendulkar, Purva, Kalpesh Krishna, Ramprasaath R. Selvaraju, and Devi Parikh (2019). "Trick or TReAT : Thematic Reinforcement for Artistic Typography." In: *Proceedings of the Tenth International Conference on Computational Creativity, Charlotte, North Carolina, USA, June 17-21, 2019*. Ed. by Kazjon Grace, Michael Cook, Dan Ventura, and Mary Lou Maher. Association for Computational Creativity (ACC), pp. 188–195. URL: http://computationalcreativity.net/iccc2019/papers/iccc19-paper-48.pdf (cit. on p. 55).
Teng, Norman Y and Sewen Sun (2002). "Grouping, simile, and oxymoron in pictures: A design-based cognitive approach." In: *Metaphor and Symbol* 17.4, pp. 295–316 (cit. on p. 82).
Terveen, Loren G (1995). "Overview of human-computer collaboration." In: *Knowledge-Based Systems* 8.2-3, pp. 67–81 (cit. on p. 47).

Torralba, Antonio and Aude Oliva (2003). "Statistics of natural image categories." In: *Network: computation in neural systems* 14.3, p. 391 (cit. on p. 64).

Tufte, Edward R. (1997). *Visual explanations: images and quantities, evidence and narrative*. Graphics Press, Cheshire, Connecticut (cit. on p. 292).

Ungerer, Friedrich and Hans-Jorg Schmid (2006). *An introduction to cognitive linguistics*. Pearson Education Limited (cit. on pp. 15–20, 32, 73–76, 78).

Ustalov, Dmitry (2012). "A text-to-picture system for russian language." In: *RuSSIR 2012* (cit. on p. 55).

Van Leeuwen, Theo (2001). "Semiotics and iconography." In: *Handbook of visual analysis*. Ed. by Carey Jewitt and Theo Van Leeuwen. Sage Publications London, pp. 92–118 (cit. on pp. 1, 25).

Van Mulken, Margot, Rob Le Pair, and Charles Forceville (2010). "The impact of perceived complexity, deviation and comprehension on the appreciation of visual metaphor in advertising across three European countries." In: *Journal of Pragmatics* 42.12, pp. 3418–3430 (cit. on p. 83).

Van Weelden, Lisanne, Alfons Maes, Joost Schilperoord, and Reinier Cozijn (2011). "The role of shape in comparing objects: How perceptual similarity may affect visual metaphor processing." In: *Metaphor and Symbol* 26.4, pp. 272–298 (cit. on p. 86).

Veale, Tony (2018). "Changing channels: divergent approaches to the creative streaming of texts." In: *Annals of Mathematics and Artificial Intelligence*, pp. 1–18 (cit. on p. 62).

Veale, Tony and Khalid Al-Najjar (2016). "Grounded for life: creative symbol-grounding for lexical invention." In: *Connection Science* 28.2, pp. 139–154 (cit. on p. 279).

Veale, Tony and F Amílcar Cardoso (2019). *Computational creativity: The philosophy and engineering of autonomously creative systems*. Springer (cit. on pp. 3, 45).

Veale, Tony and Mike Cook (2018). "Magic Carpets." In: *Twitterbots: Making Machines that Make Meaning*. Cambridge, MA: MIT Press (cit. on p. 62).

Venkatesh, Viswanath, Susan A Brown, and Hillol Bala (2013). "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems." In: *MIS quarterly* 37.1 (cit. on p. 4).

Ventura, Dan (2016). "Mere generation: Essential barometer or dated concept." In: *Proceedings of the Seventh International Conference on Computational Creativity*. Sony CSL, Paris, pp. 17–24 (cit. on p. 45).

– (2019). "Autonomous Intentionality in Computationally Creative Systems." In: *Computational creativity: The philosophy and engineering of autonomously creative systems*. Ed. by Tony Veale and F Amílcar Cardoso. Springer. Chap. 3, pp. 49–69 (cit. on p. 63).

Vernillo, Paola (2018). "The Role of the Image Schemas in the Analysis of the Semantic Variation of Action verbs. Data from IMAGACT." In: *TriCoLore (C3GI/ISD/SCORE)* (cit. on p. 20).

Vertolli, Michael (2014). "Contextual Coherence in the Visual Imagination: An Interdisciplinary Analysis." PhD thesis. Carleton University (cit. on p. 57).

Vidal, Leticia, Gastón Ares, and Sara R Jaeger (2016). "Use of emoticon and emoji in tweets for food-related emotional expression." In: *Food Quality and Preference* 49, pp. 119–128 (cit. on pp. 300, 301).

Vinhas, Adriano, Filipe Assunção, João Correia, Anikó Ekárt, and Penousal Machado (2016). "Fitness and Novelty in Evolutionary Art." In: *EvoMUSART*. Vol. 9596. Lecture Notes in Computer Science. Springer, pp. 225–240 (cit. on pp. 180, 206).

Walker, Mort (2000). *The lexicon of comicana*. iUniverse Bloomington (cit. on pp. 30, 35).

Wang, Xin, Pascal Matsakis, Lana Trick, Blair Nonnecke, and Melanie Veltman (2008). "A study on how humans describe relative positions of image objects." In: *Headway in Spatial Data Handling*. Springer, pp. 1–18 (cit. on pp. 33, 124).

Warchoł, Adam Tomasz (2018). *Conceptual Blending and the Arts: An Analysis of Michał Batory's Posters*. Cambridge Scholars Publishing (cit. on pp. 73, 76–78, 90).

Ware, Colin (2012). *Information visualization: perception for design*. Third. Morgan Kaufmann (cit. on pp. 12, 13, 29).

Wertheimer, Max (1938). "Laws of organization in perceptual forms." In: *A source book of Gestalt psychology*. Ed. by W. Ellis. London: Routledge & Kegan Paul, pp. 71–88 (cit. on p. 14).

Whale, George (2002). "Why use computers to make drawings?" In: *Computers and Art*. Ed. by Stuart Mealing. Intellect Books, pp. 17–32 (cit. on p. 2).

Wheeler, Alina (2009). *Designing Brand Identity: An Essential Guide for the Whole Branding Team*. 3rd (cit. on p. 21).

Whigham, Peter A, Colin Aldridge, and Michel de Lange (2009). "Constrained evolutionary art: Interactive flag design." In: *2009 IEEE Congress on Evolutionary Computation*. IEEE, pp. 2194–2200 (cit. on p. 253).

White, Tom (2019). "Shared Visual Abstractions." In: *arXiv preprint arXiv:1912.04217* (cit. on p. 65).

Wicke, Philipp (2017). "Ideograms as semantic primes: Emoji in computational linguistic creativity." Bachelor thesis, University of Osnabrück (cit. on pp. 145, 150, 151).

– (2021). "Computational Storytelling As An Embodied Robot Performance With Gesture And Spatial Metaphor." PhD thesis. University College Dublin (cit. on p. 45).

Wicke, Philipp and Marianna Bolognesi (2020). "Emoji-based semantic representations for abstract and concrete concepts." In: *Cognitive*

Processing. DOI: 10.1007/s10339-020-00971-x. URL: https://doi.org/10.1007/s10339-020-00971-x (cit. on p. 150).

Wicke, Philipp and João Miguel Cunha (2020). "An Approach for Text-to-Emoji." In: *Proceedings of the Eleventh International Conference on Computational Creativity* (cit. on pp. 6, 321, 329).

Wiggins, Geraint A. (2006). "A preliminary framework for description, analysis and comparison of creative systems." In: *Knowl. Based Syst.* 19.7, pp. 449–458. DOI: 10.1016/j.knosys.2006.04.009. URL: https://doi.org/10.1016/j.knosys.2006.04.009 (cit. on p. 45).

Wijeratne, Sanjaya, Lakshika Balasuriya, Amit P. Sheth, and Derek Doran (2017a). "A semantics-based measure of emoji similarity." In: *Proc. of WI-17* (cit. on pp. 145, 304).

Wijeratne, Sanjaya, Lakshika Balasuriya, Amit Sheth, and Derek Doran (2016). "EmojiNet: Building a Machine Readable Sense Inventory for Emoji." In: *8th International Conference on Social Informatics*. Springer (cit. on p. 158).

– (2017b). "EmojiNet: An Open Service and API for Emoji Sense Discovery." In: *Proceedings of ICWSM-17* (cit. on pp. 155, 158, 159, 161, 243, 257, 269, 302, 303, 330).

Wilkinson, Leland (2006). *The grammar of graphics*. Springer Science & Business Media (cit. on p. 243).

Wills, Charles, ed. (2008). *Complete Flags of the World*. DK Publishing (cit. on p. 255).

Wills, Graham and Leland Wilkinson (2010). "Autovis: automatic visualization." In: *Information Visualization* 9.1, pp. 47–69 (cit. on p. 242).

Wiseman, Sarah and Sandy JJ Gould (2018). "Repurposing Emoji for Personalised Communication: Why□ means "I love you"." In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, p. 152 (cit. on p. 147).

Wittenbrink, Craig M, Alex T Pang, and Suresh K Lodha (1996). "Glyphs for visualizing uncertainty in vector fields." In: *IEEE Trans. on Visualization and Computer Graphics* 2.3 (cit. on p. 241).

Worrall, David (2020). "Computational Designing of Sonic Morphologies." In: *Organised Sound* 25.1, pp. 15–24 (cit. on pp. 41, 42).

Wright, Glen (2011). "Your Flag's Got My Flag on it: The Union Jack and the Australian Flag." In: *Crux Australis* 98 (cit. on pp. 265, 268).

Wu, Huikai, Shuai Zheng, Junge Zhang, and Kaiqi Huang (2019). "Gp-gan: Towards realistic high-resolution image blending." In: *Proceedings of the 27th ACM international conference on multimedia*, pp. 2487–2495 (cit. on p. 66).

Xiao, Ping and Simo Linkola (2015). "Vismantic: Meaning-making with Images." In: *Proceedings of the Sixth International Conference on Computational Creativity, ICCC-15* (cit. on pp. 58, 61, 284).

Xu, Pengfei, Jianqiang Ding, Hao Zhang, and Hui Huang (2019). "Discernible image mosaic with edge-aware adaptive tiles." In: *Computational Visual Media* 5.1, p. 4 (cit. on p. 80).

Yan, Xiaoyong, Ying Fan, Zengru Di, Shlomo Havlin, and Jinshan Wu (2013). "Efficient learning strategy of Chinese characters based on network approach." In: *PloS one* 8.8 (cit. on p. 36).

Yang, Hongyi, Chengqi Xue, Xiaoying Yang, and Han Yang (2021). "Icon Generation Based on Generative Adversarial Networks." In: *Applied Sciences* 11.17. ISSN: 2076-3417. DOI: 10.3390/app11177890. URL: https://www.mdpi.com/2076-3417/11/17/7890 (cit. on p. 64).

Yannakakis, Georgios N, Antonios Liapis, and Constantine Alexopoulos (2014). "Mixed-initiative co-creativity." In: *Proceedings of the Ninth Conference on the Foundations of Digital Games* (cit. on pp. 3, 47, 205).

Ye, Keren, Narges Honarvar Nazari, James Hahn, Zaeem Hussain, Mingda Zhang, and Adriana Kovashka (2019). "Interpreting the Rhetoric of Visual Advertisements." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (cit. on pp. 89, 94–96).

Yu, Rongrong, Ning Gu, and Michael J Ostwald (2021). *Computational Design: Technology, Cognition and Environments*. CRC Press (cit. on pp. 42, 43).

Zakraoui, Jezia, Moutaz Saleh, and Jihad Al Ja'am (2019). "Text-to-picture tools, systems, and approaches: a survey." In: *Multimedia Tools and Applications* 78.16, pp. 22833–22859 (cit. on p. 56).

Zantides, Evripides et al. (2016). "Visual metaphors in communication: Intertextual semiosis and déjà vu in print advertising." In: *Revista Română de Comunicare şi Relaţii Publice* 18.3, pp. 65–74 (cit. on p. 82).

Zhang, Lingzhi, Tarmily Wen, and Jianbo Shi (2020). "Deep image blending." In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 231–240 (cit. on p. 66).

Zhao, Feng and Hao Wang (2010). "An Intelligent and Interactive Icon Creation System for Mobile Users." In: *2010 Fifth International Conference on Systems*. IEEE, pp. 1–5 (cit. on pp. 53, 54).

Zhao, Nanxuan, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson W.H. Lau, Jose Echevarria, and Zoya Bylinskii (2020). "ICONATE: Automatic Compound Icon Generation and Ideation." In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (cit. on pp. 2, 60, 274, 323).

Zhou, Bolei, Vignesh Jagadeesh, and Robinson Piramuthu (2015). "Conceptlearner: Discovering visual concepts from weakly labeled image collections." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500 (cit. on p. 280).

Zhu, Bin (2002). "Information visualization for knowledge repositories: applications and impacts." PhD thesis (cit. on p. 248).

Zhu, Jichen, Antonios Liapis, Sebastian Risi, Rafael Bidarra, and G Michael Youngblood (2018). "Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation." In: *2018 IEEE CIG*, pp. 1–8 (cit. on p. 259).

Zhu, Xiaojin, Andrew B Goldberg, Mohamed Eldawy, Charles R Dyer, and Bradley Strock (2007). “A text-to-picture synthesis system for augmenting communication.” In: *AAAI*. Vol. 7, pp. 1590–1595 (cit. on pp. 55, 56).

Zimbardo, Philip G. and Richard J. Gerrig (2002). “Perception.” In: *Foundations of cognitive psychology: core readings*. Ed. by Daniel J Levitin. MIT press. Chap. 7 (cit. on pp. 12, 29).