Imperial College London,
Department of Computing

A Constructive Theory of Automated Ideation

Flaminia Cavallo

Master of Philosophy
2014

Abstract

In this thesis we explore the field of automated artefact generation in computational creativity with the aim of proposing methods of generation of ideas with cultural value. We focus on two kinds of ideas: *fictional concepts* and *socially embedded concepts*.

For fictional concepts, we introduce a novel method based on the non-existence-conjectures made by the HR automated theory formation system. We further introduce the notion of typicality of an example with respect to a concept into HR. This leads to methods for ordering fictional concepts with respect to three measurements: novelty, vagueness and stimulation. We ran an experiment to produce thousands of definitions of fictional animals and then compared the software's evaluations of the non-fictional concepts with those obtained through a survey consulting sixty people. The results showed that two of the three measurements have a correlation with human notions.

For socially embedded concepts, we apply a typicality-based classification method, the Rational Model of Classification (RMC), to a set of data obtained from Twitter. The aim being the creation of a set of concepts that naturally associate to an initial topic. We applied the RMC to four sets of tweets, each corresponding to one of four initial topics. The result was a set

of clusters per topic, each cluster having a definition consisting of a set of words that appeared recurrently in the tweets. A survey was used to ask people to guess the topic given a set of definitions and to rate the artistic relevance of these definitions. The results showed both high association percentage and high relevance scores. A second survey was used to compare the rankings on the social impact of each of the definitions. The results obtained show a weak positive correlation between the two rankings.

Our experiments show that it is possible to automatically generate ideas with the purpose of using them for artefact generation. This is an important step for the automation of computational creativity because most of the available artefact generation systems do not explicitly undertake idea generation. Moreover, our experiments introduce new ways of using the notion of typicality and show how these uses can be integrated in both the generation and evaluation of ideas.

Acknowledgments

Thank you to my current supervisors, Krysia Broda and Alessandra Russo, for the help they gave me on a project they knew little about and for pushing me and supporting me in the decision of getting an M.Phil.

Thanks also to my previous superiors, Simon Colton, Alison Pease and John Charnley for their enthusiasm and the help they gave me in exploring my ideas.

Thanks to Alan Bundy, for encouraging me into this experience and for giving me his support throughout it.

Thanks to Amani El Kholi for keeping track and being so helpful with each change of status throughout my PhD./M.Phil.

Thanks to my examiners Geraint Wiggins and Duncan Fyfe Gillies, who made my viva an enjoyable experience. They have made many valuable contributions to the thesis.

Thanks to the old Computational Creativity Group at Imperial College London, for being weird and awesome. Robin Baumgarten, Michael Cook, Zena Hira, Ramin Ramezani, Kate Reed, Lukas Platinsky: thanks for listening to all my indecisions. You made the office a cool place to spend the day in.

Also thanks to my lunch companions, Doru Bercea, Serge Emerge, Zafiris Fountas, Jake Gillespie, Rob Harrison, Jan Jachnik, Rayane M. Jackson, Fabio Luparini, Gary Willis and especially Stefanie Glassford, for helping me escape moments of frustration and for bearing with my angry and exuberant peaks.

A final thank you to my spell checking team: Stefanie Glassford, Sabika Shah Povia, Robin Baumgarten, Ramin Ramezani, Ali Eslami, Bruno Panama, Andrew Hodgson, Rory Mulloy, Matt Mahony, Kyla Finlay, Hannah Reynolds, Alex Fernandes, Josh Allison, Nicola Hillhouse, Hassan Veldstra, Harry Cullis, Simon Hayward, Grainne Ellis, Sean Blumenfeld. You are absolute stars (I wonder who is going to spell-check this part). I hope you learned something interesting!

Declaration of Originality

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated or cited otherwise in the text. I state that this work has not been submitted for any other degree or professional qualification except as specified.

Copyright Declaration

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Research in artificial intelligence (AI) has always been focused on reality. However, since J.P. Guilford's seminal Presidential Address to the Psychological Association in 1950 and the following inclusion of creativity in the Structure of Intellect (SOI) [37], the role that ideation (or idea generation) plays within human intelligence has become relevant not only for the psychology community, but also for researchers in artificial intelligence.

Current research in Computational Creativity can be divided into three non-mutually exclusive groups, according to the general objectives of the projects in the groups: the use of software to understand human creativity; the use of software to help people produce creative work; and the use of software to manifest creative behaviours either domain independently or in a particular field. Of these, the third is the focus of this work. In particular, we are interested in the creative autonomy that can be attributed to systems that automatically produce artefacts such as paintings, games, poems and music, starting from a topic of interest, or an idea. We argue that the difficulties

expressed by the public in judging software as imaginative can be partly dependent on the fact that, while ideas are at the heart of such artefacts, they are usually introduced by the programmer.

The aim of this project is therefore to show that it is possible to implement a computational process capable of producing ideas of cultural value, as evidenced by them being embedded in artefacts. By ideas, we mean the notions that can drive a rendering process and by cultural value, we refer to the level of appreciation of a rendering by an audience.

To undertake this task, we will start by studying the formalisations underlying automated concept formation systems. Examples of projects undertaken in this direction include the development of automated theory formation software, concept blending software, analogical reasoning software and others, as described in Section 2.1. In particular, we focus on the analysis of different forms of representation that concepts take within this research. We then introduce a working definition of an idea, which is coherent with and inclusive of current AI research, but which embraces notions introduced by the psychological community in studies of the different facets of creativity, as reviewed in Section 2.4.

We then look at methods to build concepts that are coherent with our definition of ideas and discuss techniques to evaluate and interpret the ideas produced. To undertake these tasks, we make use of notions and theories from cognitive psychology. In particular, we study how the concept of *Typicality* and the concept formation theories that revolve around it can be used for the generation and creation of ideas. In the field of cognitive psychology, typicality is thought of as one of the key notions behind concept representation. Its importance was one of the main factors that led to the first criticisms of the

16

classical view [65], which argues that concepts can be represented by a set of necessary and sufficient conditions. Current cognitive theories therefore take into account the fact that exemplars can belong to a concept with a different degree of membership: this is the *typicality* of an exemplar with respect to that concept. One of the key features of typicality, is that it is dependent on experience. As a knowledge base grows and changes with the integration of new knowledge, typicality also changes. This is an aspect which is key to the creative process, where ideas are highly dependent on personal and current experiences. It is therefore in our interest to study the influence that a typicality factor might have in automated creative concept formation.

The overall scope of this project is bidirectional. On one side, we want to extend concept formation techniques to generate the kind of ideas that artists are inspired by when creating artworks. On the other side, we hope to build a bridge between cognitive concept formation techniques and automated artefact generation by allowing the former to provide inputs for the latter. As human artists speak about an inspiring "muse", we believe that our idea generation software could serve as an inspirational input to the large array of software oriented to the creation of expressive artefacts.

With this project, we do not aim to cover all aspects of creativity appearing in both psychological theories and computational techniques, or to provide a method for idea generation that can be considered exhaustive. We instead want to build an initial framework to orient and frame further research in this area and focus on the study of two methods of formation for two subsets of ideas: *fictional concepts* and *socially embedded concepts*. Fictional concepts are concepts which have no evidence in reality. Socially embedded concepts are concepts that derive from the interaction of a group of people.

Within the work described in the rest of the thesis, we demonstrate that it is possible to automatically generate ideas that could serve as a basis to the creation of artefacts, and that the notion of typicality can be successfully used to do so. We then propose a framework based on the results obtained.

The rest of this thesis is arranged as follows: in Chapter 2, we present a literature review that discusses existing work from both a computational creativity and a cognitive psychology point of view. In Chapter 3, we provide a definition and explanation for the terms *idea* and *ideation*. In Chapters 4 and 5, we describe and discuss the work we have undertaken. In particular, in Chapter 4 we discuss a method for the creation of fictional concepts, and in Chapter 5 we discuss a method for the creation of socially embedded concepts. Finally, in Chapter 6 we draw some conclusions and discuss future work.

Publications

The following publications arose from, or were based on, the work in this thesis:

- Uncertainty Modelling in Automated Concept Formation. Flaminia Cavallo, Simon Colton, and Alison Pease. In Proceedings of the Automated Reasoning Workshop 2012 [14]. This publication is based on parts of the work reported in Chapters 2, 4 and 6.

- Using Theory Formation Techniques for the Invention of Fictional Concepts. Flaminia Cavallo, Alison Pease, Jeremy Gow, and Simon Colton.

In Proceedings of the Fourth International Conference on Computational Creativity, page 176, 2013 [15]. This publication is based on the work reported in Chapter 4.

Chapter 2

Literature Review

In this thesis, we aim to provide an understanding of idea generation which takes into consideration current representations of concepts in computational creativity software and in concept formation software, but also references notions from concept representation in cognitive psychology and psychological theories on aspects of creativity. As the research in each of these areas is expansive, we focus our review on the parts that we consider of most relevance to our studies. We start by analysing some AI techniques for concept formation in Section 2.1 and for expressive rendering in Section 2.2. We then look at psychological theories of concept representation in Section 2.3 and on the different aspects of creativity in Section 2.4. Finally, we review some notions of curiosity and interestingness in Section 2.5.

2.1 Concept Formation Techniques

In psychology, the term *concept formation* usually refers to the extraction of a sets of common and relevant features from a group of items, whereas in artificial intelligence this ability is usually associated with machine learning.

Below we report a summary of those computational techniques that perform concept formation by following a discovery oriented approach either deliberately seeking novelty or deliberately reproducing some other aspects of the creative reasoning processes. These are of particular interest to our project because they constitute a large inspiration for the definition and implementation of methods to define and evaluate ideas.

2.1.1 Automated Theory Formation

Automated Theory Formation concerns the formation of interesting theories, starting with some initial knowledge and enriching it by performing inductive and deductive reasoning. In the late 1970s, Lenat developed the Automated Mathematician (AM) [44], a system which, given a large number of mathematical concepts and heuristic rules, performs interestingness-guided manipulations of these concepts in order to obtain new ones. Although AM is still important in terms of its innovational contribution, the system has been criticised for both its non-evident creative value [47,64,72] and for being strictly domain dependent [47]. Some of AM's limitations were addressed by Lenat himself in the development of the system EURISKO [46], which uses meta-heuristics to generate new heuristics as needed. EURISKO was considerably more successful and obtained satisfactory results in different fields, including VLSI chip design [48] and role-playing games [45]. However, the

system has not been widely used, probably because of its reliance upon many domain-specific rules [68].

A subsequent project in this direction lead to the development of HR [20], a system by Colton et al., that performs both concept formation and conjecture making by applying a concise set of production rules and empirical pattern matching techniques to an initial knowledge base. These rules are applied in the order dictated by an agenda, containing information on how to construct the next new concept. The production rules take as input the definition of one or two concepts and output the definition of the new concept, whose success set – the collection of all the tuples of objects which satisfy the definition – is then calculated. These sets of positive examples are then compared and hence used to formulate conjectures about the new concepts. These conjectures take the form of equivalence conjectures (when two sets of positive examples match), implication conjectures (when one set of positive examples is a subset of another), or non-existence conjectures (when a set of positive examples is empty). The conjectures are either proved by the OTTER theorem prover [51], rejected because of a counterexample found by the MACE model generator [52] or left open. HR follows a best-first non-goal-oriented search, dictated by an ordered agenda and a set of heuristic rules used to evaluate the interestingness of each concept. HR was developed to work in mathematical domains, but different projects have demonstrated the suitability of this system to work in other domains such as games [6], puzzles [19] and HR's own theories [18].

In Chapter 4, we use a modified version of the HR program for the generation of fictional concepts. In order to allow the reader to fully comprehend this chapter, we provide below the details on the HR algorithm.

HR: Constants, Concepts and Conjectures

In HR, a theory is constituted of four kind of elements: constants, concepts, conjectures and proofs.

- Constants: These are defined by a predicate of the form T(A) where A is a constant and T specifies the type of A. Examples are animal(dog) or covering(feathers).

- Concepts: These are used to express newly discovered entities and are represented by a classification rule (concept definition), the success set of such a classification rule (set of exemplars) and the classification that this rule implies (a partition on the specific subset of tuples of constants). An example is:
  $DEFINITION : Concept24(x, y) = has\_covering(x, y)$ AND $has\_milk(x)$
  $EXAMPLES : f(dog) = \{hair\}, f(bat) = \{hair\},$
  $f(dolphin) = \{none\}$ [1]
  $CLASSIFICATION : [bat, dog], [dolphin]$

- Conjectures: Conjectures are represented by an association rule (the definition of the conjecture), and other information such as its status: proved, disproved or open conjecture. An example is:
  $CONJECTURE : \forall(x)(has\_milk(x) \leftrightarrow of\_class(x, Mammal))$
  $STATUS : proved$

---

[1] Where $f(x) = \{y | Concept24(x, y)\}$

- Proofs: Proofs are represented as text output from OTTER [51] . Proofs are limited to the mathematical domain, and hence can not be applied to the above example.

HR: Production Rules

We report here some of the production rules that HR uses in order to construct new concepts: each of these rules can be applied to one (unary rules) or two (binary rules) known concepts.

- Match: equates variables in a definition.
  For instance, suppose that the old concept is: integers, a, b, c where $a * b = c$ (the concept of multiplication). Then the Match rule could be used to equate a and b and hence obtain: integers, a, c where $a * a = c$ (the concept of squares).

- Split: instantiates one or more variables in a definition.
  For instance, suppose that the old concept is: integers, a, b, c where $a * b = c$ (the concept of multiplication). Then the Split rule could be used to instantiate b to 2 and obtain: integers, a, c where $a * 2 = c$ (the concept of multiplying a number by 2).

- Exists: introduces an existential quantifier over one or more variables in a definition.
  For instance, suppose that the old concept is (as in the example above): a, c where $a * 2 = c$ (the concept of multiplying a number by 2). Then the Exist rule could be used to invent the concept of all integers for which there exists a number that gives this integer when multiplied by 2: this is the concept of even numbers.

- Size: counts the number of distinct tuples of constants in the success set of a definition.

  For instance, suppose that the old concept is: integers, a, b, where the number of divisors of $a = b$. Then the Size rule could be used to count the number of divisors each integer has.

- Compose: uses conjugation to compose the clauses of two concepts into a new definition.

  For instance, given the concept of the number of divisors of an integer and the concept of even numbers, the Compose rule could be used to invent the concept of integers with an even number of divisors.

- Negate: negates certain clauses in one of the definitions.

  For instance, given the concept of even numbers the Negate rule could be used to invent the concept of odd numbers.

HR: Interestingness Measures

HR performs concept formation by following a best-first search, dictated by an ordered agenda containing information on how to construct the next new concept. Each item in the agenda represents an instruction on what production rule to apply to which existing concept(s) and with which parameters. The agenda is ordered with respect to the interestingness of the concepts for development. These values are calculated as a weighted sum (where the weights are provided by the user) of the following measures:

- Applicability: the proportion of constants found in the success set of a concept.

- Comprehensibility: the reciprocal of the number of production steps that went into building the concept.

- Parsimony: the reciprocal of the number of elements in the success set multiplied by the arity of the definition.

- Variety: the number of different classes in the classification.

- Development Steps: how many production rules a definition has been involved in, which gives an indication of how much it has been developed.

- Productivity: the proportion of theory formation steps the concept has been used in and that have successfully produced a new concept.

- Novelty: 1 minus the proportion of other concepts in the theory which achieve the same classification.

- Parents: the average interestingness of the parents of the concept (this can only be used in conjugation with other measures).

- Children: the average interestingness of the children of the concept (this can only be used in conjugation with other measures).

- Proof Difficulty: the average difficulty (as assessed by OTTER) of proved theorems about the concept.

- Invariance: if the user specifies a desired classification, this is the proportion of all pairs of objects which should be classified together and that are classified together.

- Discrimination: if the user specifies a desired classification, this is the proportion of all pairs of objects which should be classified as different and that are classified as different.

2.1.2 Conceptual Combination

In cognitive psychology, one of the most studied general purpose processes of creative thinking is conceptual combination. Conceptual combination involves the merging of previously separate ideas, concepts, visual forms etc. [78]. In computational creativity, research in this direction aims towards the construction of models that reason in different domains, inclusive of a transition mechanism that allows for the transferred knowledge to make sense in the new context [59]. A distinction is usually made between systems that perform conceptual combination to enforce convergent thinking and those that use it to enforce divergent thinking. In the first case, conceptual combination is used to enrich or modify a given data structure; in the second case, conceptual combination is used to create completely new and independent data structures: this is the case of conceptual blending systems.

The theory behind conceptual blending was initially formalised by Gilles Fauconnier and Mark Turner [32]. They describe conceptual blending as a procedure which, given two initial mental spaces, generates a third one, called the *blend*, by obeying a selected structure mapping. The new domain will partially maintain the structure provided by the input domains, but will also add its own independent structure.

The Conceptual Blending process can be divided into three stages [32]:

- Composition: this process involves the projection of the knowledge representing each input into the blended space. This process might involve the union of some of the features of the inputs [36].

- Completion: during this process, frames encoded into the knowledge base are used to fill out the details of a pattern in the blend. The completion phase often results in the emergence of a new structure [36].

- Elaboration - during this process, the structure in the blend is elaborated through some cognitive work performed with the blend [36].

One of the most recent conceptual blending models is *Divago*, developed by Francisco Pereira [59]. Divago uses both reasoning techniques and genetic algorithms to produce its blends. It has been applied to both visual and linguistic domains, giving satisfactory results in terms of its creative contribution [59]. One of the most illustrative examples for which Divago has been used is the automated generation of concepts such as Pegasus [58]. To complete this task, Divago is given two concept maps: one for the concept 'Horse' and another of the concept 'Bird'. A concept map consists of a set of first order logic statements that characterise the concept it represents. The system is also provided with *frames* (used to describe specific composite concepts, such as 'new ability') and *integrity constraints* (used to maintain soundness within concepts). These frames and integrity constraints are used to manipulate both of the two initial concept maps, by integrating one into the other. The result consists of a set of new concept maps, each describing a blend between the two initial concepts; in this case, the concept of a horse and the one of a bird (and hence the set of resulting concept maps can be thought of as a set of different versions of Pegasus). More detailed informa-

28

tion about the algorithm used in Divago can be found in Section 4.4, where
we compare it to our fictional concept creation method.

2.1.3 Analogical Reasoning

Analogical reasoning is another process which has received attention from
AI researchers and cognitive scientists [31, 34], and has strong connections to
computational creativity [56, 77]. This can be described as the application or
projection of structured knowledge from a familiar domain to a novel and less
familiar one [78]. A frequently mentioned example of analogical reasoning is
Rutherford's explanation of the hydrogen atom through the comparison with
the solar system.

From a computational point of view, one of the most cited models of analog-
ical reasoning is Falkenhainer et al.'s Structure-Mapping Engine (SME) [31],
based on Gentner's structure-mapping theory of analogy [34]. SME takes as
input a base concept and a target concept and, by building a match between
the structures of the two, it outputs an interpretation of the comparison.
Such an interpretation consists of three parts: the correspondences between
the structures of the base and target concepts, a set of possible inferences
about the target concept which can be implied from the matches between
the structures, and a score for the quality of the match. For example, the
two structures represented in Figure 2.1 can be used by SME to interpret the
analogy between the solar system and an atom. The algorithm first looks
for all the possible mappings between the two structures. In Figure 2.1, for
example, the system creates various possible mappings: (i) one that matches
the nucleus to the sun, and the planets to the electrons; (ii) one that matches
the mass of the sun to the mass of the nucleus; and so on. A *structural support*

score is then given to each of these mappings, proportional to the number of possible inferences that can be derived from it. In the example above, mapping (i) would imply that in an atom, as in the solar system, the difference in masses between the nucleus and the electrons causes the electrons to revolve around the nucleus.
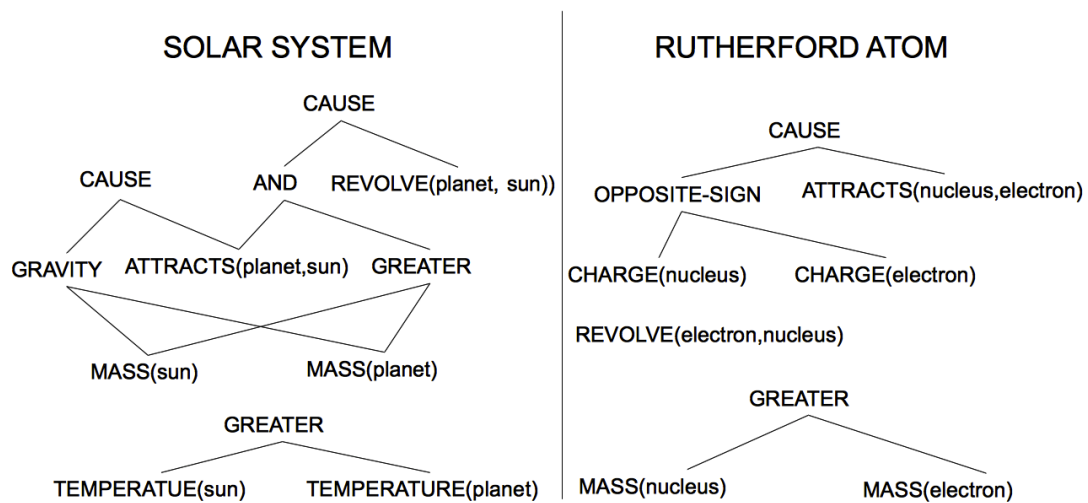
Figure 2.1: Simplified 'Solar System' and 'Rutheford Atom' structures used in SME.

2.1.4   Inductive Logic Programming

Out of the many of methods proposed in the field of Machine Learning, we believe that Inductive Logic Programming (ILP) is the one of most relevant to our project because of its descriptive representation and its suitability for learning relational predicates [66]. Moreover, ILP has connections with non-goal-oriented reasoning (in descriptive ILP, explained below) and with probabilistic reasoning (in probabilistic ILP, explained below), which are both central to our project.

Given a knowledge base and a set of observations, ILP uses inductive reasoning in order to advance a set of hypotheses in the form of logic programs. We describe three different approaches to ILP: predictive ILP, descriptive ILP and probabilistic ILP.

In predictive ILP, given a knowledge base and a set of observations in the form of positive and negative exemplars, the goal is to find a hypothesis that covers all of the positive and none of the negative exemplars. This is achieved by starting with a general rule and specialising it gradually so that it is consistent with the observations. Hence predictive ILP can be used when a classification for a set of exemplars is known and the goal is to find a rule to explain this classification [66].

Probabilistic ILP integrates probabilistic reasoning with the first order representations and methods used in predictive ILP. In this case, the clauses in the set of observations are annotated with probability values and the task of the algorithm is to find a hypothesis that maximises the likelihood of these observations [29].

In descriptive ILP, the goal is to find, given a knowledge base as a set of predicates, a hypothesis that explains the data. Hence descriptive ILP systems require no classification labels (no sets of positive and negative exemplars). Hypotheses are generated through the formation of new predicates. Descriptive ILP can then be described as a non-goal-oriented learning system, as it provides a method to enrich a knowledge base without the need for specifications on what the user is looking for.

2.2 Automated Expression of Ideas

We describe here the branch of Computational Creativity research focusing on the automation of systems able to produce creative artefacts in a particular domain, such as paintings, poems, games, music etc. The amount of research undertaken in this direction is large, and the techniques used are generally dependent on the domain of interest.

Many projects have been carried out in the linguistic field, through the development of software for the generation of poems [16, 35], jokes [9, 10, 73] and stories [13, 53, 75]. There has also been research in the music field such as [27, 30, 80, 82], the visual field e.g. [21, 50], and in other fields such as games [25] and cooking [54].

Generally speaking, we can divide the projects in this area into two main groups: (i) those that focus on the application and evaluation of a particular production method on different sets of data, where the methods could be either provided by the programmer or automatically learned through the analysis of human-generated work, and (ii) those that focus on the search for possible approaches that can be applied in order to obtain a pre-defined result.

In the first case, the general objective is to understand how and when the application of a given technique could be considered creative. Examples of systems here include the painting system AARON [50], whose initial goal was to discover "What are the minimum conditions under which a set of marks functions as an image?" [50], or those systems that automatically learn the writing/music/painting style of a particular artist in order to produce more work in that style. In the second case, the objective is reversed: the role of

the system is to interpret and embrace a given idea, and to find an adequate way to express it through a form of art. We call this second process *automated expression of ideas*.

In this thesis, we focus on such systems, and we argue that the difficulties expressed by the public in judging these as imaginative can be partly dependent on the fact that, while ideas are at the heart of such creative artefacts, they are usually introduced by the programmer.

Some progress in this direction has been made by systems that automatically capture ideas from a given resource, such as a news article or a photograph. We have summarised three of these systems below. Note that whilst these systems remove the role of humans in choosing and interpreting an idea, the ideas are still provided by people (i.e the journalists or the photographer etc.).

2.2.1 The Painting Fool

The Painting Fool [21] is a program that automatically generates paintings without the need of user intervention: the program tries to reproduce not only the characteristics of a good painting ("by applying a set of instructions" [17]) but the whole painting process ("by providing its own set of instructions" [17]). In order to be considered an artist in its own right, The Painting Fool was constructed with the aim of being attributed a notion of intent, and was hence developed so that its paintings are generated not from human given directions, but from data extracted from the internet.

An example comes from automated collage generation [43], which is achieved through the retrieval of news article, the extraction of keywords and the

retrieval of images from Flickr using these keywords. This data is used for the construction of input files specifying which images to annotate and extract colour segments from, how to arrange the segments in an overall collage, and what natural media to simulate when painting the segments to produce the final piece [43].

Cook and Colton point out that this system somewhat removes the role of human decision making: humans have in fact little control over (a) what news story will be chosen for a collage (b) what keywords will be extracted (c) what images will be retrieved or (d) how the collage will be rendered [24]. However, given that the data used is still human provided, the intentionality behind the production of the collages can be attributed not only to the software but in an equal amount also to the articles' writers, to the photographers, to the programmer, the interpreting public and the people that tagged images in Flickr [24].

2.2.2 ANGELINA

ANGELINA [25] is a system that can automatically design videogames through the use of co-operative co-evolution techniques adopted to evolve simple platform games. ANGELINA uses a variety of social and other media to design a game. In one project, the starting point is a Guardian newspaper article. This gives ANGELINA a set of keywords and topics that the article covers (like 'Afghanistan', 'Finance', 'David Cameron' or 'The Arab Spring'). These can then be used to power searches on other websites for more media and data. The keyword data goes through a process of refinement where ANGELINA uses sources like Wikipedia to add factual or classifying information to keywords [26].

For example, Afghanistan can be labeled as a country, and David Cameron as a person. Once the keywords are refined, ANGELINA then starts to look for more subjective data about them – photos, sound effects, and opinions. For instance, knowing that a keyword refers to a person, ANGELINA queries Twitter with "$\langle name \rangle is$" and looks at the words attributed to that person. This gives an estimate for the public opinion towards these people, which is then used to alter the searches that ANGELINA performs for photos. If a keyword is identified as a country or location, ANGELINA uses Flickr's geolocation API to find photographs taken in that area and uses them as background images in a video-game. Some data use is more abstract - the program processes the article's raw body text and runs it through a sentiment analysis routine. If the article is depressing or sad, ANGELINA looks for melancholic music to play during the game; or upbeat music if the article is happy.

2.2.3   Poetry Generation

The use of internet resources for the construction of expressive artefacts has also been applied to poem generation [23]. Colton at al. propose a system which uses templates to construct poems according to given constraints on rhyme, meter, stress, sentiment, word frequency and word similarity [23]. The software uses newspapers to construct a mood for each day, to select an article on which to base a poem on and to choose a template for the poem. Subsequently, the program generates an aesthetic based on relevance to the article, lyricism, sentiment and flamboyancy [23], and it searches for an instantiation of the template which maximises the aesthetic. Finally, the program provides a commentary for the whole process to add value to the

creative act.

2.3 Concept Representation in Cognitive Psychology

Given our objective of building a system which forms ideas for human-oriented purposes, we decided to look at cognitive theories of concept representation in order to formalise a data structure for our ideas, and at cognitive theories of concept formation in order to inspire our methods of idea formation. Throughout the corresponding literature, the term *category* usually refers to a class of things: a subset of a set of entities, grouped together with respect to some reasonable criteria. The word *concept* refers to the mental representation of a category [55]. The debate on what is the correct representation lies at the base of the differences between theories of categorization. Below we report a summary of these theories and how they evolved through time.

2.3.1 From the Classical View to the Exemplar and Prototype View

Until the 1970s, concepts were regarded as being mentally represented by a definition: this school of thought is usually referred to as the *classical view*. In the classical view, an item is said to belong to a category if it meets a set of necessary and sufficient conditions: the concept definition. This view hence implies the law of the excluded middle, according to which an item

either belongs or doesn't belong to a category. It also follows a mathematical logic approach, which although is very simple and elegant, is unfortunately not plausible for most real life examples: it is in fact not only hard to come up with a definition for most of the world's categories but, more importantly, it is essential to make a distinction between categories' members [55]. For example, the concept of 'Dog' could be defined as 'a four-legged barking mammal with fur'. But how can we classify a dog that has lost a leg? What about a toy dog animal?

A complete theory must therefore take into account both the *typicality* of items (by recognising that some items are more typical of a category than others) and the existence of "in-between categories" cases. One last criticism of the classical view referred to its transitive properties in categorization, which are not always consistent in real case scenarios: if A belongs to B, and B belongs to C, A does not need to belong to C [55]. For example, a 'car chair' is a kind of 'chair', and a 'chair' is a kind of 'furniture', but this doesn't imply that a 'car chair' is 'furniture'.

Following Roch's first critique of the classical view in 1970 [65], the recognition of these problems led to the formation of two main streams of thought: the *prototype view* and the *exemplar view*.

The prototype view states that concepts are mentally represented by a list of the features that are usually found among the categories' members. Features do not need to be consistent with one another, and can also be contradictory. For example, the prototype of the concept of 'dog' could be equal to the following list of features: {white, black, brown, barks, hasTail, etc.}. A weight is given to each feature according to its frequency. When we classify

a new object, its features are compared with those from the feature list, and the object is assigned to the category with which it shares most features. According to this view, we can then hypothesise the existence of a member for each category which best represents it: the concept's *prototype*. This does not need to be a concrete example, and is represented as a list of the most common features. In a more complete version of the prototype view, features are organised into dimensions, so that each feature belongs to one, and only one, of these dimensions. Additional weight is given to each dimension according to how determinant it is to the concept definition. This second representation not only provides more information about the category, but also introduces constraints which stop the formulation of incoherent concepts.

The exemplar view states that a concept is mentally represented by a set of the category's specific and remembered instances (the *exemplars*) [71]. The exemplar view hence assumes that people store in memory information about every instance of a stimulus, along with information about its category membership. For example, the concept of dog might be stored as the following set: {Bobby, Pongo, Lucky, etc.}, where Bobby, Pongo and Lucky are dogs. Categorization is then achieved by computing the similarity of an item with each of the remembered exemplars in a category.

Both the exemplar and prototype views therefore imply that some members are better examples of a concept than others, and hence that some members are more *typical* of a category than others. Also, both views imply that categories' boundaries are fuzzy, as items might be members of two or more categories.

2.3.2   Further Work in Concept Representation

The analysis of the differences between the exemplar and prototype views led many researchers to realise that these two models could correspond to special (extreme) cases of a more generic model. This realisation lead to the formulation of a range of clustering-based algorithms such as the VAMP model [76], the SUSTAIN model [49] and the ALCOVE model [42]. One of the first, and arguably the most used, of these models is Anderson's Rational Model of Classification (RMC) [5], which proposes an algorithm to build concepts from stimuli in a dynamic, probabilistic and order-dependent way. The system loops through these phases:

- Given a set of concepts, the *knowledge set* (note that this could be equal to the empty set) and a new item, compute this item's degree of membership with respect to each concept in the knowledge set. Also compute the likelihood that the item belongs to a new, not yet defined concept.

- Take the highest likelihood: this will determine whether the new exemplar belongs to the corresponding concept or a whole new concept (and hence that the item is the first encountered member of a previously unknown concept).

- Modify the selected concept definition by integrating the features of the new item.

Models such as the RMC are important not only because they propose a method of unification between the prototype and exemplar schools of thought (by recognising them to be extreme cases of a generic model), but also because

they regard both concept representations and the categorization process as dynamic and flexible entities. That is, a concept could change as more instances are classified, and the classification process is dependent on the order that stimuli are encountered. We will expand on these features in Section 6.1.2.

From the above discussion, it is evident that a lot of progress has been made in this field since Roch's first critique in 1970 [65] of the classical view of categorization. However, while research in cognitive psychology continues to develop on formulation of new concept representation theories, current computational creativity software still rely on a strictly definitional approach. In this project, we adopt notions suggested from the cognitive psychology literature, such as the use of prototypes, typicality measures and a dynamic classification approach for the definition and construction of ideas. In particular, the notion of typicality is heavily used in Chapter 4, while the RMC is used in Chapter 5.

2.4 Psychological Theories of Creativity

The amount of research by cognitive psychologists devoted to the study of different aspects of creativity and to the formulation of theories of creativity continues to grow. This is of no surprise given the fact that despite the recognition of its importance, discrepancies still exist between opinions on what creativity actually means. In the following sections we report a summary of the theories of creativity that are relevant to this project.

2.4.1 Boden's Three Ways of Creativity

In [12], Boden identifies three different types of creativity: *combinational creativity*, *exploratory creativity* and *transformational creativity*. To explain these, she first introduces the concept of a *conceptual space*. A conceptual space is the set of all possible items which satisfy a predefined set of rules or constraints (noting that such a set could be infinite). Examples could include all the possible ways of making a painting, or all the possible moves in a game of chess. In Boden's own words a conceptual space is 'any disciplined way of thinking that's familiar to (and valued by) a certain social group' [12].

The first kind of creativity that Boden identifies is *combinational creativity*. It involves making unfamiliar comparisons of familiar parts of distinct conceptual spaces (or distinct parts of the same conceptual space), where such comparisons are considered creative if not seen before. For example, this kind of creativity would include making a comparison between an atom and the solar system.

The other two types of creativity introduced by Boden, namely *exploratory creativity* and *transformational creativity*, involve the exploration or the transformation of a conceptual space. Exploratory creativity occurs when a part of a search space which has never been visited is taken into consideration. For example, in a game of chess many moves are possible, but not all of them might have been used. Then the use of a new move would be considered an example of exploratory creativity. As another example, in visual art there are infinite possibilities of 'lines of color' that can be drawn on a canvas, but only some of them have been used in the past. A new combination of 'lines of color' is another example of exploratory creativity. Transformational creativity instead occurs when the rules or the constraints of a conceptual space

are changed or expanded. Boden describes it as follows: 'the deepest cases of creativity involve someone's thinking something which, with respect to the conceptual spaces in their minds, they couldn't have thought before. The supposedly impossible idea can come about only if the creator changes the pre-existing style in some way'. In the two examples used above, transformational creativity would occur if a new rule which maintains the consistency and soundness of the previous rules is added to a game of chess, or if a new way to draw on a canvas is used.

In [81], Wiggins formalises Boden's conceptual space, exploratory and trasformational creativity. Wiggins identifies two set of rules (details can be found in [81]) which define a conceptual space: the constraints of the space, which determine whether an item belongs to the space, and the exploratory rules of the space, which determine how the items in a conceptual space are searched for. The author then points out that this implies the existence of two types of tranformational creativity: one involves transforming the constraints of a conceptual space, hence allowing the inclusion of new items into it; the other involves the transformation of the exploratory rules of a conceptual space, hence allowing the discovery of items that already belonged to the conceptual space, but could not be found before. We refer back to Boden's ways of creativity and Wiggins' formalization in Section 6.1.2.

2.4.2   The four P's of Creativity

Different theoretical studies on creativity can be framed with respect to the facets that they give prominence to. Traditionally, theories of creativity can be defined as focusing on one or more of the following aspects: process, product, person and place [60]. These are usually referred to as the *four P's*

of creativity.

Theories that address the *Creative Process* are the ones that aim to comprehend the cognitive mechanisms that occur when someone is engaged in creative thinking or creative activity [41]. Some recurrent themes among these theories are the study of commonalities and differences between creative thinking and non-creative thinking, the rules of conscious versus unconscious process and the contribution of stochastic processes versus more controlled and guided processes [41].

Theories that focus on creative products aim at the analysis of concrete results. These theories usually provide quantitative and objective methods for the measurements of the value of an artefact, focusing on its novelty and usefulness. It is, however, important to remember that a lot of researchers argue that the study of a product is ineffective in terms of understanding creativity if little can be said about the process leading to it or the creator's personality [41].

Another branch of studies on creativity address the *creative person.* These theories generally try to analyse personality traits that could be indicative of a creative person, independently of the domain or in a specific domain.

Finally, theories that focus on the *creative place* (or *creative press*) analyse the settings in which creative acts take place. Recurrent themes in this area of research include the interactions between people, or between a person and the environment. General agreement among these theories validate the fact that creativity tends to manifest itself when there are opportunities for explorations and independent work, and when originality is supported. An important theory that can be classified into this group is Csikszentmihalyi's

system view of creativity [28]. He argues that the processes essential to creativity are to be found in the interactions between individuals and society. He hence identifies three important components of creative systems: the individual, the domain (the cultural component) and the field (the social or interactive component) and argues that each of these components is essential in determining creativity.

We refer back to the four Ps of creativity in Section 6.1.1, where we propose a mapping between three of the Ps and a proposed framework.

2.4.3 Categories of Creative Magnitude

In describing theories of creativity, it is important to distinguish between levels of creative magnitude. One of the most common distinctions is between *little-c creativity* and *big-C Creativity* [41]. Big-C Creativity [61] refers to indubitable examples of creativity, like Picasso's paintings, Einstein's relativity theory and Mozart's music. Big-C creativity's works are the ones that make major contributions to the development of a field [41]. Little-c creativity refers to everyday creativity [61]: little-c creativity works might consist of novel approaches to tackle a problem or of small discoveries which are interesting but not of high importance for a domain. Berghetto and Kaufman [38] subsequently introduce two additional levels: mini-c creativity and pro-c creativity. *Mini-c creativity* was introduced to divide the subjective and objective aspects of *little-c creativity* [41]: mini-c creativity refers to creativity at a personal level, including aspects such as mental or emotional creative changes. Pro-c creativity was instead introduced to define the fuzzy area that lies between *little-c* and *big-C creativity* [41]. It refers to professional-level creators (like professional artists) who do not have eminent

44

status, but who are beyond little-c creators in knowledge, motivation and performance [78].

A similar distinction has been proposed by Boden [12], who suggested a classification aimed at the description of the behaviour of both software and humans. Boden's proposition is based on who perceives a creative product as such. When something is perceived as creative from a personal prospective, and hence is novel just to the creative person, we are referring to psychological-creativity, or p-creativity. When something is new and useful to a community or the of whole humanity, we refer to historical-creativity, or h-creativity.

2.5 Interestingness and Curiosity

Other areas of research that are strictly connected to our project are the ones that study and propose theories of interestingness and curiosity. Below we have summarised some the relevant work undertaken in these areas. We refer to the notions introduced below in Section 4.2, where we propose some measurements of interestingness based on typicality.

2.5.1 Notions of Interestingness

One of the most used ways to classify interestingness, first suggested by Silberschatz and Tuzhilin [70], is the distinction between *subjective* and *objective* interestingness. Objective interestingness focuses on the evaluation

of an item based exclusively on its properties and characteristics. Subjective interestingness instead compares properties of an item with the beliefs and the knowledge of a person.

Both Schmidhuber [69] and Csikszentmihalyi [28] underline the importance that the subjective aspects of interestingness have over the objective ones, stating that it must only be evaluated in terms of an observer's current knowledge and computational abilities.

According to Silberschatz and Tuzhilin [70], the two main characteristics of subjective interestingness are:

- Unexpectedness: a measurement inversely proportional the predictability of a result or event

- Actionability: a measurement for the number of actions that an agent could undertake as a consequence of a discovery.

In [68] interestingness is evaluated through the use of the Wundt Curve as a function that plots interest with respect to novelty. According to this theory, interestingness can be considered to be a special case of hedonic value: a measurement for the pleasure associated with heightened states of learning. Saunders hence modelled the relationship between the hedonic value (pleasure) and novelty (distance to reality) using a non-linear fuction called the Wundt Curve, shown in Figure 2.2 [7].

The Wundt curve has since been used in many models of computational creativity. It's maximum value is located in a region located relatively close to the y-axis. Saunders points out that this can be interpreted as the fact that
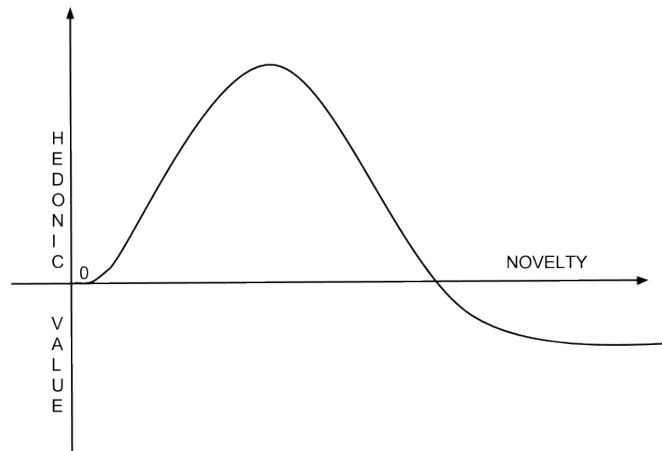
Figure 2.2: The Wundt Curve: a hedonic function used to calculate interest. The y-axis measures the hedonic value: a measurement for the pleasure associated with heightened states of learning; the x-axis measures novelty: a measure of the distance to reality. The maximum value, located in a region close to the y-axis, can be interpreted as the fact that the most interesting experiences are those that are "similar-yet-different" to those that have been experienced previously.

the most interesting experiences are those that are "similar-yet-different" to those that have been experienced previously [68].

Colton compares the measures of interestingness that have been used in some mathematical discovery systems [22]. He identifies five types of commonly used measures: Novelty (whether a concept or conjecture is new with respect to a knowledge base), Surprisingness (whether a concept or conjecture is predictable), Applicability (whether the concept or conjectures apply to a large amount of data in the knowledge base), Comprehensibility and Complexity (whether a concept or conjecture is simple enough to be understood) and Utility (whether a concept or conjecture can be used for further goals).

47

2.5.2 Ritchie's Criteria

In [62], G. Ritchie proposes eighteen criteria to asses some behaviors of a computer program which might be indicative of creative potential. These criteria aim to evaluate the creativity of a program by measuring some properties of the outputs that such a system produces. The outputs (i.e. a painting, a verbal joke) are called *basic items*, and they are considered independently of whether they are consider a successful output or unsuccessful results.

The criteria that Ritchie proposed consist of some calculation involving what he identifies as the primitive aspects of basic items: their *value* and their *typicality*. *Value* is described as a measurement of the extent to which the produced item is a high quality example of his genre [62] (i.e. 'To which extent is output a a good painting?'). *Typicality* is instead described as a measurement of the extent to which the produced item is an example of the artefact class in question (i.e. 'To which extent can output a be classified as a painting?'). Both value (val) and typicality (typ) are expressed as a mapping from the basic item to the set $[0, 1]$ (by using fuzzy sets). In order to understand the criteria, Richie makes use of following additional definitions: the *Inspiring Set I* is the subset of the available basic items which drove the creative program computation. 'It could be all the relevant artefacts known to the program designer, or items which the program is designed to replicate, or a knowledge base of known examples which drives the computation within the program' [62]. The set of basic items that a program produced is instead represented by the letter R. $T_{\alpha,1}(R)$ is the subset of R consisting of items which have typ higher than a chosen constant α. Similarly, $V_{\gamma,1}(R)$ is the subset of R consisting of items which have val higher than a chosen constant γ.

Given the above definitions, we report the formulas of Ritchie's criteria in the Table 2.1.

2.5.3 Curiosity

Berlyne defines curiosity as "a form of motivation that promotes exploratory behaviour to learn more about a source of uncertainty, such as a novel stimulus, with the goal of acquiring sufficient knowledge to reduce the uncertainty" [8]. Berlyne proposed to divide curiosity into two types: curiosity driven by diversive exploration, and curiosity driven by specific exploration. In the case of diversive exploration, a person is under-stimulated and hence seeks arousal from the environment. In the case of specific exploration, a person is over-stimulated and tries to reduce their arousal by exploring a particular situation in order to reduce uncertainty. We can then hypothesise a parallelism between these two kinds of curiosity and the two main creative processes of convergent versus divergent thinking. That is, diversive exploration stimulates divergent thinking, while specific exploration stimulates convergent thinking.

2.6 Summary

In the sections above we have reviewed some of the key studies and theories related to creativity from both an artificial intelligence and psychological prospective. Within the rest of the thesis, we refer to this work for different

Formalization
$AV(typ, R) > \theta$, for suitable θ
$Ratio(T_{\alpha,1}(R), R) > \theta$ for suitable θ, α
$AV(val, R) > \theta$, for suitable θ
$Ratio(V_{\gamma,1}(R), R) > \theta$ for suitable θ, γ
$Ratio(V_{\gamma,1}(R) \cap T_{\alpha,1}(R), T_{\alpha,1}(R)) > \theta$ for suitable θ, α, γ
$Ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), R) > \theta$ for suitable θ, β, γ
$Ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), T_{0,\beta}(R)) > \theta$ for suitable θ, β, γ
$Ratio(V_{\gamma,1}(R) \cap T_{0,\beta}(R), V_{\gamma,1}(R)) > \theta$ for suitable $\theta, \alpha, \beta, \gamma$
$Ratio(I \cap R, I) > \theta$ for suitable θ
$1 - Ratio(I \cap R, R) > \theta$ for suitable θ
$AV(typ, (R - I)) > \theta$ for suitable θ
$AV(val, (R - I)) > \theta$ for suitable θ
$Ratio(T_{\alpha,1}(R - I), R) > \theta$ for suitable θ, α
$Ratio(V_{\gamma,1}(R - I), R) > \theta$ for suitable θ, γ
$Ratio(T_{\alpha,1}(R - I), (R - I)) > \theta$ for suitable θ, α
$Ratio(V_{\gamma,1}(R - I), (R - I)) > \theta$ for suitable θ, γ
$Ratio(V_{\gamma,1}(R - I) \cap T_{\alpha,1}(R - I), (R - I)) > \theta$ for suitable θ, α, γ
$Ratio(V_{\gamma,1}(R - I) \cap T_{0,\beta}(R - I), (R - I)) > \theta$ for suitable θ, β, γ

Table 2.1: Ritchie's Criteria: formulas to evaluate the creativity of a basic item

reasons. In particular, we use the work summarised above for the following purposes:

- To identify trends and gaps within current computational creativity research. The material reviewed in Section 2.2 on current computational creativity methods (and in particular artifact generation) is utilised for this purpose, as explained in Chapter 3.

- To actively use the algorithms proposed. This is the case for HR, reviewed in Section 2.1.1 and used for the experiments in Chapter 4, and for the RMC, reviewed in Section 2.3.2 and used in the experiments in Chapter 5.

- To contextualize and justify some of the methods and conclusions reported in the following chapters. Sections 2.4, 2.5 and 2.3 were included for such purpose.

- To provide the reader with knowledge on alternative methodologies which have been used in the field. Section 2.1 has been included for such purpose, and a comparison between methodologies can be found in Section 4.4.

In the chapters that follow, we make use of the above purposes to guide the explanation and review of our studies.

Chapter 3

Ideation

"One could say that a man can inject an idea into the machine, and that it will respond to a certain extent and then drop into quiescence, like a piano string struck by a hammer. Another simile would be an atomic pile of less than critical size: an injected idea is to correspond to a neutron entering the pile from without. Each such neutron will cause a certain disturbance which eventually dies away. If, however, the size of the pile is sufficiently increased, tire disturbance caused by such an incoming neutron will very likely go on and on increasing until the whole pile is destroyed. Is there a corresponding phenomenon for minds, and is there one for machines? There does seem to be one for the human mind. The majority of them seem to be "subcritical," i.e., to correspond in this analogy to piles of subcritical size. An idea presented to such a mind will on average give rise to less than one idea in reply. A smallish proportion are supercritical. An idea presented to such a mind that may give rise to a whole

theory consisting of secondary, tertiary and more remote ideas.
Animals' minds seem to be very definitely subcritical. Adhering
to this analogy we ask: Can a machine be made to be supercrit-
ical?"

A. M. Turing [74]

3.1   A Working Definition of Ideas

In this thesis, we will propose a framework to form and evaluate ideas. Be-
fore doing so, we will provide an informal explanation and working definition
of this term. We argue that one of the limitations of current creative sys-
tems oriented to the generation of artefacts is the lack of automation on the
creation of an initial idea upon which the rendering process is then based.
For example, generally an automated painting system needs to be instructed
with something like: 'make a painting about love' by the programmer/user.
In some of the more advanced systems, such as those described in Section
2.3, the idea might be directly extracted from a piece of text, such as a news-
paper article. In these cases, the responsibility of *picking a topic* shifts from
the programmer/user to the system. However, ultimately the idea is still
generated by a human: in the example above by the journalist that wrote
the article.

In this project, we underline the necessity for a system for the automatic
generation and creation of ideas. These ideas *may or may not be consistent
with reality*, but instead need to be interesting because of their *cultural value*
and, as Turing states in the quote above, because they "may give rise to
a whole theory consisting of secondary, tertiary and more remote ideas"

[74]. We hence argue that the difficulties expressed by the public in judging software as imaginative can be partly dependent on the lack of this automated step: while ideas are at the heart of such creative artefacts, they are usually introduced by the programmer.

With these uses in mind, we provide the following working definition:

> An idea is a concept which can be used to guide the generation of an artefact and which can be evaluated in terms of the impact that this artefact has on the public.

Note that this definition is intentionally general, to allow room for extensions and different implementations. In this thesis we will however restrict ourselves to what we consider two of the basic forms of ideas: fictional concepts (Chapter 4) and socially embedded concepts (Chapter 5). By *fictional concept* we refer to, as explained above, concepts that are not consistent with reality. By *socially embedded concepts* we refer to concepts that derive from the interaction between a group of people.

The use of these ideas is evident if we look at the current Computational Creativity systems oriented towards the creation of programs for the production of artefacts, like those reported in Section 2.2: an automated poem generator [23] might compose a poem about the idea of the atrocity of war and an automated painting generator like The Painting Fool [17] might produce a picture about the idea that men could fly. These are examples of ideas that we aim to generate.

In the chapters that follow, we propose and analyse two methods for the generation and evaluation of ideas, one for fictional concepts (Chapter 4) and one for socially embedded concepts (Chapter 5).

Chapter 4

Using Theory Formation Techniques for the Invention of Fictional Concepts

Research in Artificial Intelligence has always been largely focused on reasoning about data and concepts which have a basis in reality. As a consequence, concepts and conjectures are *generated* and *evaluated* primarily in terms of their truth with respect to a given knowledge base. For instance, in machine learning, learned concepts are tested for predictive accuracy against a test set of real world examples. As underlined in previous chapters, in Computational Creativity research, much progress has been made towards the automated generation of artefacts (painting, poems, stories, music and so on). When this task is performed by people, it might start with the conception of an idea, upon which the artefact is then based. Often these ideas consist of concepts which have no evidence in reality. For example, a novelist could write a book centered on the question 'What if horses could fly?' (e.g.,

Pegasus), or a singer could write a song starting from the question 'What if there were no countries?' (e.g., John Lennon's Imagine). However, in Computational Creativity, the automated generation and evaluation of such fictional concepts for creativity purposes is still largely unexplored.

The importance of evaluating concepts independently of their truth value has been highlighted by some cognitive science research, as reviewed in 2.5. Some of the notions that often appear in the cognitive science and psychology literature are those of *novelty*, *actionability*, *unexpectedness* and *vagueness*. Novelty is used to calculate the distance between a concept and a knowledge base. As reviewed in Section 2.5, in [68], interestingness is evaluated through the use of the Wundt Curve [7], a function that plots hedonistic values with respect to novelty. As also reviewed in Section 2.5, actionability is used to evaluate the number of actions or thoughts that an agent could undertake as a consequence of a discovery, while unexpectedness is a measurement inversely proportional to the predictability of a result or event. Finally, vagueness is referred to as the difficulty of making a precise decision. Several measurements have been proposed in the literature for the calculation of this value, particularly using fuzzy sets [40].

The importance of *generating* concepts which describe contexts outside of reality was underlined by Boden when she proposed her classification of creative activity. As previously discussed, Boden identifies 'three ways of creativity' [11]: *combinational creativity*, *exploratory creativity* and *transformational creativity*. While combinational creativity involves making unfamiliar combinations of familiar ideas [11], and exploratory creativity requires the discovery of unknown areas of a search space, transformational creativity involves the modification of a search space by breaking its boundaries.

One reading of this could therefore be the creation of concepts that are not supported by a given knowledge base; we refer to these as fictional concepts herein. Conceptual blending [33] (reviewed in Section 2.1.2) offers clear methods for generating fictional concepts, and we return to this later, specifically with reference to the Divago system which implemented aspects of conceptual blending theory [59].

In this chapter, we propose a new approach to the formation and evaluation of fictional concepts. Our method is based on the use of the HR automated theory formation system [20] (reviewed in Section 2.1.1), and on cognitive science notions of concept representation (reviewed in Section 2.3). In particular, we explore how the notion of *typicality* can improve and extend HR's concept formation techniques.

In the following sections, we discuss the methods and results obtained by introducing typicality values into HR. We argue that such typicality measures can be used to evaluate and understand fictional concepts. In particular, we propose calculations for three measures which might sensibly be linked to the level of *novelty*, *vagueness* and *stimulation* associated with a fictional concept. We generated definitions of fictional animals by applying our method to a knowledge base of animals and we report the results. We then compare the software's estimate of novelty, vagueness and stimulation with data obtained through a questionnaire asking sixty people to evaluate some concepts with the same measures in mind. The results were then used to test whether there is a correlation between our measurements and the usual (human) understanding of the terms novelty, vagueness and stimulation. We then compare this approach and the well established methods of conceptual blending. Finally, we draw some conclusions and discuss some further

directions.

4.1 Using HR to Generate Fictional Concepts

We are interested in the generation and evaluation of concepts for which it is not possible to find an exemplar in the knowledge base that completely meets the concept's definition. Throughout this chapter, we use the term *fictional concepts* to refer to this kind of concept. We use the HR system for the generation of such fictional concepts. To do so, after it has formed a theory of concepts and conjectures in a domain, we look at all the non-existence conjectures that it has generated. These are based on the concepts that HR constructs which have an empty success set. Hence, the concepts that lie at the base of these conjectures are fictional with respect to the knowledge base given to HR as background information. For example, from the non-existence conjecture:

$$\nexists(x)(Reptile(x) \,\&\, HasWings(x))$$

we extract the fictional concept:

$$C_0(x) = Reptile(x) \,\&\, HasWings(x)$$

To see whether typicality values can be used for the evaluation of these fictional concepts, we have introduced this notion into HR. Typicality values are obtained by calculating the degree of membership of each user-given constant (i.e., animals in the above example) with respect to every fictional concept which specialises the concept of the type of object under investigation (which is the concept of being an animal in this case). This is done by looking

at the proportion of predicates in a concept definition that are satisfied by each constant. Hence, for each constant a_j and for each fictional concept C_i in the theory, we will have $Typicality(a_j, C_i) = t$, where $0 \leq t < 1$. For example, for the concept definition:

$$C_1(x) = Mammal(x) \,\&\, HasWings(x)$$
$$\&\, LivesIn(x, Water)$$

the typicality values for the constants in the set $\{Lizard, Dog, Dolphin, Bat\}$ are as follows:

$$Typicality(Lizard, C_1) = 0;$$
$$Typicality(Dog, C_1) = 0.\overline{3};$$
$$Typicality(Dolphin, C_1) = 0.\overline{6};$$
$$Typicality(Bat, C_1) = 0.\overline{6};$$

We see that the constant 'Dolphin' has typicality of $0.\overline{6}$ with respect to C_1 because a dolphin is a mammal which lives in water but which doesn't have wings – hence it satisfies two of the three predicates ($\approx 66.6\%$) in the definition of C_1.

We use a simple measure to calculate typicality, and we are aware that it could be improved in multiple ways as explained in 4.5. However, it is sufficient to demonstrate the point of this experiment, and we'll hence leave improvements to future studies. It is important to note that for each fictional concept C there are at least n constants $a_1, ..., a_n$ such that $\forall j$, $0 < Typicality(a_j, C) < 1$, where n is the number of predicates in the concept

60

definition. This is because HR requires at least one exemplar for each predicate in the initial knowledge base. We refer to these as the *atypical exemplars* of fictional concept C, and we denote this set of constants as $atyp(C)$. The atypical exemplars of C have typicality bigger than zero because they partly belong to C, and less than one because the concept is fictional, and hence by definition it doesn't have any real life examples. The number of atypical exemplars of a fictional concept is always more than or equal to the number of predicates in the concept definition, because fictional concepts originate from the manipulation of non-fictional concepts, and hence, – given a well formed knowledge base – each predicate in a fictional concept definition will correspond to a non-fictional concept with at least one element in its success set.

4.2   Evaluating Concepts Based on Typicality

We explain here how typicality can be used to evaluate fictional concepts along three axes which we claim can be sensibly used to estimate how people will assess such concepts in terms of vagueness, novelty and stimulation respectively. This claim is tested experimentally in the next section. To define the measures for a fictional concept C produced as above, we use E to represent the set of constants (examples) in the theory, e.g., animals, and we use NF to denote the set of non-fictional concepts produced alongside the fictional ones. We use $|C|$ to denote the number of conjunct predicates in the clausal definition of concept C. We further re-use $atyp(C)$ to denote the set of atypical exemplars of C and the *Typicality* measure we introduced above. It should be noted that the proposed methods of evaluation of fic-

tional concepts have not been included into the HR program to guide concept formation, but rather applied after theory formation has occurred.

4.2.1 Using Atypical Exemplars

Our first measure, M_V, of fictional concept C, is suggested as an estimate of the *vagueness* of C. It calculates the proportion of constants which are atypical exemplars of C, factored by the size of the clausal definition of C, as follows:

$$M_V(C) = \frac{|atyp(C)|}{|E| * |C|}$$

As previously mentioned, vagueness is a measurement that has been widely studied in the context of fuzzy sets. Klir [40] emphasises the difference between this measurement and the one of *ambiguity*, and underlines how vagueness should be used to refer to the difficulty of making a precise decision. While several more sophisticated measurements have been proposed in the literature, as explained in [40], we chose the above straightforward counting method, as this is consistent with the requirement that if concept C_a is intuitively perceived as more vague than concept C_b, then $M_V(C_a) > M_V(C_b)$. To see this, suppose we have the following two concepts:

$$C_1(x) = Animal(x) \ \& \ has(x, Wings)$$
$$C_2(x) = Reptile(x) \ \& \ has(x, Wings)$$

In this case, we can intuitively say that an animal with wings is more vague than a reptile with wings, because for the first concept, we have a larger choice of animals than for the second. In terms of typicality, this can be interpreted as the fact that C_1 has a larger number of atypical exemplars than C_2, and it follows that $M_V(C_1) > M_V(C_2)$.

4.2.2 Using Average Typicality

Our second measure, M_N, of fictional concept C, is suggested as an estimate of the *novelty* of C. It calculates the complement of the average typicality of the atypical exemplars of C, as follows:

$$M_N(C) = 1 - \frac{1}{|atyp(C)|} \left(\sum_{a \in E} Typicality(a, C) \right)$$

Novelty is a term largely discussed in the literature, and can be attached to several meanings and perspectives. In our case, we interpret novelty as a measurement of distance to the real world, as inferred in previous work in computational creativity research, such as [68]. As an example of this measure, given the concepts:

$$C_1(x) = Bear(x) \& Furniture(x) \& Has(x, Wings)$$
$$C_2(x) = Bear(x) \& Furniture(x) \& Brown(x)$$

then, in a domain where all the constants are either exclusively bears or furniture (but not both), and assuming that all the bears and all the furniture are brown, we calculate:

$$M_N(C_1) = 0.\overline{6}$$
$$M_N(C_2) = 0.\overline{3}$$

This is because for C_1, all exemplars will satisfy just one of the three clauses ($\frac{1}{3}$) in the definition, hence this will be their average typicality, and C_1 will score $1 - \frac{1}{3} = 0.\overline{6}$ for M_N. In contrast, all exemplars will satisfy two out of the three clauses in C_2, and hence it scores $0.\overline{3}$ for M_N. Hence we can say that C_1 is more distant from reality, and hence more novel, than C_2. Consistent with the literature, and in particular with the Wundt Curve (which compares novelty with the hedonic value), we assume that the most interesting

concepts have an average typicality close to 0.5. Note that this implies that fictional concepts whose definition contains two conjuncts are always moderately interesting in terms of novelty, as their average typicality is always equal to 0.5.

4.2.3 Using Non-Fictional Concepts

Our final measure, M_S, of fictional concept C is suggested as an estimate of the *stimulation* that C might elicit when audiences are exposed to it (i.e., the amount of thought it provokes). It is calculated as the weighted sum of all the non-fictional concepts, r, in NF that HR formulates for which their success set, denoted $ss(r)$, has a non-empty intersection with $atyp(C)$. The weights are calculated as the sum of the typicalities over $atyp(C)$ with respect to C. $M_S(C)$ is calculated as follows:

$$M_S(C) = \sum_{r \in NF} \left(\sum_{a \in atyp(C) \cap ss(r)} Typicality(a, C) \right)$$

This calculation is motivated by Ward's path-of-least-resistance model [79]. This states that when people approach the task of developing a new idea for a particular domain, they tend to retrieve basic level exemplars from that domain and select one or more of those retrieved instances as a starting point for their own creation. Having done so, they project most of the stored properties of those retrieved instances onto the novel ideas they are developing. As an example, the fictional concept:

$$C_1(x) = Horse(x) \,\&\, Has(x, Wings)$$

could lead to the following questions: Is it a mammal? Can humans ride it?

Animals(x): Horse, Frog, Eagle, Shark, Bee, ...

BodyPart(x): Leg, Hoof, Trunk, Tail, Eye, ...

Ability(x): Flying, Swimming, Hunting, Carrying, Food, ...

Existence(x,y)	
Frog	Forest
Frog	Grass
...	

Pw(x,y)	
Horse	Leg
Horse	Hoof
Parrot	Beak
...	

isA(x,y)	
Horse	Mammal
Eagle	Bird
...	

HasAbility(x,y)	
Horse	Run
Horse	Carry
Owl	Fly
...	

Class(x): Mammal, Fish, Reptile, Bird, ...

Place: Ocean, Arctic, Forest, Grass, ...

Purpose(x): Walk, Grab, Eat, See, ...

HasPurpose(x,y)	
Leg	Walk
Eye	See
Mouth	Eat
...	

Figure 4.1: Details from the knowledge base for animals.

Does it live in a farm? Does it fly? Does it lay eggs? Each of these questions can be derived from the corresponding HR generated concepts which have in their success set a large number of the atypical exemplars of C_1.

4.3 Experimental Results

To evaluate our approach, we started with a knowledge base of animals, based on similar inputs to those used for the conceptual blending system Divago [59], which is described in the next section. The concept map for a horse was taken from [58] and reapplied to each animal from a list of 69 animals reported in the National Geographic Kids website[1]. The relations were maintained when relevant, and extended when necessary according to the Generalized Upper Model hierarchy, as instructed in [59]. Figure 4.1 illustrates a small part of the information we provided as background knowledge for HR to form a theory with.

To generate fictional concepts with HR, we used a random-search setup and

---

[1]kids.nationalgeographic.co.uk/kids/animals/creaturefeature

ran the system for 100,000 steps, which took several hours. We limited the HR system to use only the *compose*, *exists* and *split* production rules, as described earlier in Section 2.1.1. Extracting them from non-existence conjectures, the system produced 4623 fictional concepts, which were then automatically ranked in terms of their M_V, M_N and M_S values, as described above. From each of the ranked lists, a sub-list of 14 fictional concepts was created. The fictional concepts were taken at regular intervals so that they were evenly distributed numerically over the sub-lists, from highest scoring to lowest scoring. For the M_N sub-list, all the fictional concepts with two clauses in the definition were first filtered out. For the M_V and M_S sub-lists, all the fictional concepts with more than two clauses in the definition were filtered out instead. The resulting sub-lists are given in tables 6.1, 6.2 and 6.3 respectively.

We performed a survey of sixty people who were shown these lists and asked to rank them from 1 to 14 with respect to their own interpretations of the fictional concepts and their values. The aim of the survey was to verify how measurements M_V, M_N and M_S described above correlate with respect to common (human) understanding of vagueness, novelty and stimulation respectively. The survey was composed of four parts. The first three parts asked people to rank the three sets of 14 concepts in terms of vagueness, novelty and stimulation. We didn't include an explanation of our interpretation of these words in the questions, to encourage participants to use their own understanding of the three terms. The fourth part of the survey asked for a qualitative written definition of each of the three criteria of evaluation: vagueness, novelty and stimulation. Tables 6.1, 6.2 and 6.3 in the Appendix report the three sub-lists of fictional concepts and the ranking (1 to 14) that our software assigned to them, along with the rankings obtained from the

66

survey.

In order to establish whether our ranking and the survey rankings are correlated, we calculated Pearson's correlation, r, between the system's ranking and an aggregated ranking. The aggregated ranking was calculated by ordering the fictional concepts 1 to 14, according to the mean rank from the participants. We then calculated the respective 95% Confidence Intervals (CI) and p-values, using the alternative hypothesis that the correlations are greater than zero. We obtained the following results (quoted to 3 decimal places):

$$M_V/\text{vagueness: } r = 0.552, p = 0.020, 95\% \text{ CI} = [0.124, 1]$$

$$M_N/\text{novelty: } r = 0.697, p = 0.003, 95\% \text{ CI} = [0.350, 1]$$

$$M_S/\text{stimulation: } r = -0.029, p = 0.059, 95\% \text{ CI} = [-0.481, 1]$$

We can therefore conclude that there is strong and highly statistically significant correlation between the software rankings given by M_N and the survey rankings for novelty. We have similarly found a significant and moderate correlation with the survey rankings for M_V. Hence it appears that the novelty and vagueness measurements we suggested offer sensible calculations for the general understanding of these two terms for fictional concepts.

We found no correlation between the survey rankings for the stimulation value and the software measure M_S. This could be due to two reasons. Firstly, looking at the general descriptions of the word 'stimulating' given by people in the last section of the survey, they present a broader range of meanings than the words 'novel' or 'vague'. Moreover, these meanings are often very distant from the interpretation of the term 'stimulation' that we

Figure 4.2: Word clouds: vagueness, novelty and stimulation.

used in deriving the M_S measure. In Figure 4.2, we present word clouds obtained from the definitions that people in the survey gave of the words vagueness, novelty and stimulation respectively. We can see that the word cloud for vagueness includes words such as 'description', 'unclear' and 'difficult' as might be expected, and the word cloud for novelty includes words such as 'different', 'unusual' and 'original', also as expected. However, the word-cloud for 'stimulation' includes words such as 'emotion', 'exciting' and 'imagination'. This suggests a second reason that could explain the lack of correlation: our measure M_S lacks factors to estimate emotions and surprisingness elements.

To explore the question of stimulation further, we looked at another measure of fictional concepts which might give us a handle on this property. Table 4.1 in the Appendix portrays the non-fiction concepts found (during the experimental session with HR described above) to have examples overlapping with the atypical exemplars of this fictional concept: $C_p(A) =$

$isa(A, equine), pw(A, wings)$ [noting that $pw(A, X)$ means that animal A has a body (p)art (w)ith aspect X]. These non-fiction concepts comprised the subset of NF that was used to calculate $M_S(C_p)$. The non-fiction concepts overlapping with C_p are given along with a calculation which was intended to capture an essence of C_p as the *likelihood* of additional features being true of the fictional animals described by C_p. The calculation takes the sum of the typicalities of the atypical exemplars of the fictional concept which are also true of the non-fiction concept. We see that it is more likely for the winged horse to have feathers than to have claws, as pw(A,feathers) scores 10, while pw(A,claws) scores just 1. These likelihood scores could be used at the heart of new measures. For instance, we can hypothesise that the inverse of average likelihood over all the associated non-fiction concepts might give an indication of how thinking about C_p could lead to less likely, more imaginative and possibly more stimulating real world concepts.

4.4   A Comparison with Conceptual Blending

We compare our system to the well-established conceptual blending technique, as this technique performs fictional concept formation and evaluation, as defined above. We therefore present a comparison of our system with Divago [59], which is a conceptual blending system implemented on the basis of the theory presented in [33]. It applies the notions suggested by this theory in order to combine two concepts into a stable solution called a *blend*. Blends are novel concepts that derive from the knowledge introduced via the inputs, but which also acquire an emerging structure of their own [59].

Divago has been successfully tested in both visual and linguistic domains [59].

CONCEPT: isanimal(A,horse), pw(A,wing)	
Non-fictional concept	Likelihood of predicate to characterise fictional concept
isa(A,bird)	6.5
isa(A,bug)	3.0
isa(A,mammal)	1.0
pw(A,lung)	8.5
pw(A,mane)	0.5
pw(A,tail)	7.0
pw(A,claws)	1.0
pw(A,teeth)	1.0
pw(A,eye)	10.5
pw(A,legs)	10.5
pw(A,fur)	1.0
pw(A,feathers)	10.0
pw(A,beak)	10.0
pw(A,hoof)	0.5
pw(A,claw)	5.5
existence(A,mountain)	2.5
isa(A,bug)	3.0
isa(A,bird)	6.5
isa(A,mammal)	1.0
hasAbility(A,carry)	1.0
hasAbility(A,hunt)	1.5
hasAbility(A,flying)	8.0

Table 4.1: Non-fiction concepts with success sets overlapping with atypical exemplars of the given concept, along with their actionability.

It is comprised of six different modules: the *knowledge base*, the *mapper*, the *blender*, the *factory*, the *constraints module* and the *elaboration module*. The knowledge base contains the following elements: *concept maps* that are used to define concepts through a net of relations; *rules* that are used to explain inherent causalities; *frames* that provide a language for abstract or composite concepts; *integrity constraints* that are used to assess the consistency of a concept; and *instances* that are optional sets of examples of the concepts. The mapper takes two random or user selected concepts and builds a structural alignment between the two respective concepts maps. It then passes the resulting mapping to the blender, which produces a set of projections. Each element is projected either to itself, to nothing, to its *counterpart* (the elements it was aligned with by the mapper), or to a compound of itself and its counterpart. The blender therefore implicitly defines all possible blends

that constitute the search space for the factory.

The factory consists of a genetic algorithm used to search for the blend that is evaluated as the most satisfactory by the constraints module. The algorithm uses three reproduction rules: asexual-reproduction, where the blend is copied; crossover, where two blends exchange part of their lists of projections; and mutation, where a random change in one of the projections in a blend is applied. The factory interacts both with the elaboration module and the constraints module. The elaboration module is used to complete each blend by applying context-dependent knowledge provided by the rules in the knowledge base. The constraints module is used for the evaluation of each blend. It does this by measuring its compatibility with the frames, integrity constraints, and a user-specified goal [59].

The first high-level difference between Divago and our system derives from the motivations behind their implementations. Divago was constructed to test the cognitive plausibility of a computational theory of conceptual blending, and hence their aims were to construct complete and stable concepts, i.e., the blends. Details of the system's reasoning process, used for the formation and elaboration of such concepts, are therefore presented in the final output. Our system was instead constructed to generate fictional ideas of value. These are concise concepts which are deliberately left in a simple and ambiguous form. The aim is in fact to find the concepts that stimulate the highest amount of thought and interest in an audience. The system's reasoning process is hence hidden from the outputs, and used only for evaluation purposes.

In the following paragraphs, we describe the parallels between Divago's modules and the different components of our system. In doing so, we identify the

71

consequences of using each methodology. The first comparison that can be made is between the structures of the user-provided knowledge bases. In HR, the knowledge base is used only to define a set of concepts. It is hence equivalent in functionality to Divago's concept maps. The rules, frames and integrity constraints that need to be user-specified in Divago, are instead automatically learned in HR. They take the form of conjectures, non-fictional concepts and function specifications respectively. On one hand, this implies that HR has a greater degree of autonomy. On the other hand, HR is more prone to errors, as the constructed conjectures, non-fictional concepts and functions may not be relevant for the construction of fictional concepts.

For example, given an appropriate knowledge base, HR could construct the concept of an animal being amphibious, which is defined as an animal that lives in water and lives on earth. The same frame can be manually defined and used in Divago. However, HR will simultaneously construct other similar concepts. For example, the concept of animals that live in water and are red; or the concept of animals that live on earth and have four legs. If we assume that these concepts could be used for the evaluation of fictional concepts, then there is currently no way to differentiate between them in terms of the relevance they might have on the definition of a fictional concept (i.e., the system couldn't itself determine that an amphibian is more relevant than a water-living red animal). Moreover, HR is not capable of constructing all the rules, frames and constraints that Divago uses, but we believe that a similar functionality could be achieved through the use of typicality-based exemplar membership.

Despite the evident differences between their internal mechanisms, we can make a comparison between the blends produced by Divago's mapper and

blender modules, and HR's non-existence conjectures. The first observation regards the range of the potential outputs. For HR, we only consider the concepts that are empirically known to be fictional. Divago's blends could instead be fictional, non-fictional, or exact copies of the two initial inputs. Moreover, Divago focuses only on one of the possible bijections between the elements in the concept maps. Pereira recognises that this restriction narrows the creative potential of the system [59, p.117]. HR is instead able to consider all possible structural alignments. Furthermore, Divago works on the blend of two randomly selected or user specified concepts, while HR can consider multiple concepts at once.

A component to develop and elaborate on HR's fictional concepts is still missing from our system. In order add this component, one could take inspiration from Divago's factory and elaboration modules, while also taking into consideration the typicality values discussed above. However, as explained before, in our case this reasoning module would be used to calculate the *potential reasoning* that can originate from a fictional concept. In Divago, the factory and elaboration modules are instead used for the completion of a blend. Finally, Divago's constraints module can be compared with measures M_V, M_N and M_S introduced above. Divago's constraints module aims to evaluate a completed blend, while our system rates fictional concepts. Nevertheless, a correspondence between the evaluation methods can be noted. For example, the *topology* constraint used in Divago measures the novelty of a blend, like the M_N measure for fictional concepts investigated above, and the *integration* constraint used in Divago measures how well-defined a blend is, which is similar to the M_V measurement that we have found is positively correlated with vagueness.

73

4.5   Conclusions and Further Directions

We have proposed a method for generating and evaluating fictional concepts, using the HR theory formation system enhanced with typicality values. With the experiments we have conducted, we have shown that it is possible to create fictional concepts by using this process and that it is possible to meaningfully order the fictional concepts in terms of interestingness-oriented measurements. We have compared the automatically achieved evaluations with a ranking obtained through the analysis of a survey consulting sixty people. This showed that our M_V and M_N measures are correlated positively with common understandings of vagueness and novelty respectively. We also compared our approach to the one based on conceptual blending in the Divago system, which placed our work in context and highlighted comparisons which may inform future implementations.

The experiment above indicates that our system is capable of creating fictional concepts that could be of interest to an audience. Moreover, this ideation process could be used at the heart of more sophisticated artefact generation systems, e.g., for poems or stories.

As previously discussed, the methods used to rank such fictional concepts have been shown to be useful, but also present some issues. Further research could therefore look into methods to refine the current approach and implement new measures to estimate the interestingness of fictional concepts. To do so, one could take inspiration from the notions analysed in [22] and used in the HR system, and modify them as appropriate. One could also look at other measurements suggested and used in Computational Creativity literature, such as Ritchie's criteria [62], reviewed in Section 2.5.2. These,

for example, could be used to assess the novelty of a fictional concept with respect to other fictional concepts.

Moreover, our measurement of typicality could be refined. To do so one could take inspiration from the theories proposed in cognitive science on the evaluation of the prototype theory and the weighting of category features. Each feature could be given a value called *salience*, used to indicate how important it is for the concept's definition. The salience values will then be used to calculate the typicality values with more accuracy.

Ultimately, it could be possible to introduce a notion of the *distortion of reality*. This measurement could serve to calculate how many real world constraints a fictional concept breaks. A measurement for this notion could be derived from two methods for the calculation of values related to it. The first method is introduced in [57] and is based on the number of conjectures that each atypical exemplar of a fictional concept breaks. The second method is based on the scale of the distortion that an ontology would be subject to in order to include a fictional concept.

Given that the above measurements are strictly dependent on the knowledge base used, it would be interesting to study how the construction of such knowledge base could influences the fictional concepts generated from it. This will be discussed further in Section 6.1.1. Finally, it could also be possible to implement further methods for reasoning with fictional concepts. These methods could be used to estimate actionability; for the elaboration of fictional concepts; and for potential renderings of ideas in cultural artefacts such as poems and stories. One could also study how the different methods of measurement could be related to a rendering choice and vice versa. For example, non-vague concepts could be suitable for paintings, while actionable

concepts might be more suitable for storytelling.

Chapter 5

Using Concept Formation Techniques for the Invention of Socially Embedded Concepts

This chapter describes the process and analyses the results in applying a recognized concept formation method to posts in Twitter, with the aim of inventing concepts which are considered *socially embedded* by a sample of people.

By *socially embedded*, we mean that such concepts need to describe a word/fact the same way an interactive group of people might do. Hence, they need to be represented not by the application of a dictionary definition but instead by a description which underlines the most memorable and/or interesting associations this word/fact relates to. For example, take the word 'grandmother'. The dictionary would define it as 'the mother of one's father or mother' [4]. However, if you ask a person to represent this concept artis-

tically, they would instead rely on its relation to other concepts such us 'welcoming house', 'warm cakes', 'life stories'.

In order for our concepts to be socially embedded, we believe that it is important to start with a knowledge base which reflects public beliefs and opinions. Since Twitter [1] constitutes a large and easily accessible source of such data we have decided to use it as a source for these experiments.

In order to capture the relational aspects of these concepts (to other concepts), we have decided to apply cognitive concept formation techniques that, as described in Section 2.3, strongly rely on such assumptions. Studies from cognitive psychology in fact agree that concepts are fuzzy, dynamic and experience-dependent entities [55], and should be represented as such. Recurrent ideas within these theories, such as the use of typicality and of prototypes, are key elements for our reasoning process, as they result from long-proven and discussed studies on how concepts relate to each other. For the above reasons, we have decided to adopt a well established cognitive-based concept formation method, Anderson's Rational Model of Classification (RMC) [5] (reviewed in Section 2.3.2), for our purpose.

In Section 5.2, we verify the significance of our method by proposing a comparison between the concepts automatically generated with our system from 5 words, and the results obtained from a survey that asked a sample of people to describe social associations to the same words.

The automatic creation of socially embedded concepts constitutes a relevant step forward in terms of automation for current creative systems. As noted in Section 2.3.2, existing artefact generation systems often combine logic-based representations of some aspects of the world with rules for manipulating and

representing these representations. These representations are usually built-in manually. The objective of our study is to show that concept formation techniques can be utilized to automatically generate such representations, so that they can be used in further studies on how our automatically generated concepts and existing artefact generation systems could be combined in order to develop a first fully automated creative system.

In the sections that follow, we describe how the RMC was applied to data retrieved from Twitter (Section 5.1), we present and analyse the results in comparing the output to some survey-retrieved data (Section 5.2), and we draw some conclusions and further directions (Section 5.3).

5.1 Applying the RMC to Data from Twitter

The aim of applying Anderson's Rational Model of Classification (RMC) [5] to a set of tweets is to cluster these tweets depending on the words appearing in them. Each tweet is considered to be an *exemplar* of the initial topic. Each of the words occurring in a tweet is considered to be a *feature* of an exemplar. The final clusters will hence contain a set of exemplar (tweets), each having themselves a set of features (words). The features appearing the most in a cluster will then be considered to be the *definition* of that cluster.

In order to obtain this definition, we first need to obtain a clean set of data. To do so, first a topic is decided. This is set to be a word or a combination of words. Examples could be 'War', 'Love' or 'London Olympics'. Then, a set of the most recent 1500 tweets containing these word(s) is retrieved. Such tweets are downloaded from Twitter using an external program, twitter4j [2]. To avoid trivial results, the most popular words according to the Kilgariff

database of 208,000 word frequencies [39], such as 'a', 'the', etc., and the initial topic are filtered out of each tweet. All HTML links and @ references are also filtered out from each tweet. The above procedure results in a set of 1500 clean tweets related to the initial topic. Each word appearing in any of these tweets is considered to be a *feature* of the initial topic. For example, below is a list of the cleaned up tweets that we obtained after following the above steps for the topic 'London Olympics' [1]:

- Five Travel Tips Frugal London Trip Fox Business

- Online gambling firm Betfair expects smash records

- Fans buy equestrian eventing tickets global ticket market cheap rate

- Three sessions football Wembley

- Going Robbo testimonial failed qualify

Please note that these lists could be further processed, for example, by using stemming. However, this goes beyond the scope of this thesis and we will therefore leave such improvements to future studies. The RMC method is then applied to find clusters of related exemplars (tweets). The most common features (words) per cluster will constitute the definition of this cluster. Each of these definitions is then considered to be a socially embedded concept related to the initial topic. The RMC uses a flexible representation that can interpolate between exemplars and sets of features, adding new clusters to the representation as required. When a new exemplar is analysed, this can be assigned to a pre-existing cluster, or to a new cluster on its own [67].

---

[1]These experiments were conducted in summer 2012, when they London Olympics were a current event.

The RMC is the applied by looping through the obtained set of exemplars (tweets) as follows:

Set $P_t = \{c_1, .., c_{n_t}\}$ to be a partition of t exemplars into n_t clusters, where n_t is not pre-defined and where t denotes the number of steps the algorithm has been through. Each step corresponds with the analysis of an additional exemplar. For example, if $t = 5$ then we can assume that the algorithm has considered and partitioned into n_5 clusters the first 5 exemplars, where $1 \leq n_5 \leq 5$, and it's now trying to classify the 6th exemplar. Let F_{c_i} be the set of features of cluster c_i, where $1 \leq i \leq n_t$, and $F_{e_{t+1}}$ be the set of features for the newly considered exemplar. Then the posterior probability that a new exemplar e_{t+1} was generated from cluster c_i is calculated using Bayes' Theorem as follows:

$$P(c_i|F_{e_{t+1}}) = \frac{P(F_{e_{t+1}}|c_i)P(c_i)}{\sum_i P(F_{e_{t+1}}|c_i)P(c_i)}$$

where $P(F_{e_{t+1}}|c_i)$ is the probability that the new exemplar belongs to cluster c_i given the exemplar's set of features $F_{e_{t+1}}$ and $P(c_i)$ is the prior probability that the new exemplar was generated from cluster c_i. Note that given the looping nature of the algorithm, both of these probabilities are dependent on the clusters assignments for the previous exemplars. They are calculated as follows:

$$P(c_i) = \frac{Kn_i}{(1 - K) + f_i}$$

where n_i is the number of exemplars that have been assigned to cluster c_i, f_i is the number of features of cluster c_i and K is a constant that Anderson calls the *coupling probability* [67], set to be equal to 0.8, and:

81

$$P(F_{e_{t+1}}|c_i) = \frac{\sum_{j\ in\ F_{e_{t+1}}} V(j, c_i)}{\sum_{j\ in\ F_{e_{t+1}}} V(j)}$$

where $V(j)$ is the number of times a feature j in $F_{e_{t+1}}$ has been encountered in all the previous exemplars, and $V(j, c_i)$ is the number of times a features j in $F_{e_{t+1}} \cap F_{c_i}$ has been encountered in previous exemplars.

The posterior probability that a new exemplar was generated from an entirely new cluster c_{t+1} is calculated in exactly the same way, with the exception of $P(c_i)$ which is calculated as follows:

$$P(c_{t+1}) = \frac{1 - K}{(1 - k) + f_{t+1}}$$

where f_{t+1} is the number of features of cluster c_{t+1} and K is the *coupling probability*.

Given the above formula, each exemplar e_i is considered in turn and is assigned to the cluster for which the corresponding probability is equal to:

$$Max(P(c_1|F_{e_i}), ..., P(c_{n_t}|F_{e_i}), P(c_{n_t+1}|F_{e_i}))$$

Once every exemplar has been assigned to a cluster, the definition of a cluster is equal to the set:

$$Def(c_i) = \{j in F_{c_i} | V(j, c_i) >= \frac{\sum_{k\ in\ F_1,..,F_t} V(k)}{20}\}$$

Below we show some of the clusters obtained for the topic "Changes in football":

CLUSTER c_1:

$Def(c_1) = \{$ "nothing" $\}$

Exemplars in $c_1 = \{$ "nothing", "nothing the season past was the most entertaining throughout all the divisions amazing", ... $\}$

CLUSTER c_2:

$Def(c_2) = \{$ "calling", "americans", "soccer" $\}$

Exemplars in $c_2 = \{$ "Americans calling it soccer", "people who call it f***ing Soccer", ... $\}$

CLUSTER c_3:

$Def(c_3) = \{$ "technology", "goal", "line" $\}$

Exemplars in $c_3 = \{$ ""GOAL LINE TECHNOLOGY!", "Cheaper ticket prices and goal line technology", ... $\}$

For the above example, we obtained 32 clusters. Other examples include a cluster about racism, one about Manchester United, one about diving, etc.

5.2 Experimental Results

In order to verify whether our resulting clusters' definitions can be considered to represent socially embedded concepts related to a pre-defined topic, we have constructed a survey and analysed the responses from 50 people. The survey was constructed in order to analyse the results obtained from running our program on the following four topics: 'Grandmother', 'London Olympics', 'Love' and 'War'. These topic were chosen in order to cover different kinds of socially embedded concepts. 'Grandmother' was chosen as an example of a topic than anyone can relate to, 'London Olympics' as an example of a currently relevant event[2], 'Love' as an example of a sentiment and 'War' as an example of a word with high emotional impact. The survey focused on gathering people's mental association to the above topics, and was divided into two sub-surveys, each given to 25 people, as follows:

- Sub-survey (*i*) asked people to guess a topic given the set of words in the definitions of the top ten clusters automatically created for each of the above topics. The survey allowed three different guesses. Additionally, the survey asked people to provide a vote from one to ten for each of the guesses on how well the definitions represent artistically the topic, where one corresponds to an artistically strong link, and ten corresponds to a week artistic link.

- Sub-survey (*ii*) asked people to rank from one to ten, ten different word associations related to each of the initial topics. The associations were to be ranked on their social impact, where one corresponds to the association with higher social impact, and ten corresponds to the

---

[2]At the time of this research

Topic	ρ	μ	σ
Grandmother	80%	8.25	1.33
London Olympics	48%	5.67	0.49
War	100%	8.6	1.53
Love	96%	8.2	0.76

Table 5.1: Results of sub-survey (i): the percentage ρ of people that guessed the initial topic given the set of automatically constructed definitions and the mean value μ and the standard deviation σ of the votes obtained for each correct guess.

association with lower social impact. We didn't include an explanation of our interpretation of 'social impact' to encourage participants to use their own understanding of this term. The associations provided in the survey were gathered from the results obtained by running our algorithm on the topics listed above. We used the most popular tweet for each of the top ten clusters.

The aim of sub-survey (i) is to study whether a group of people can relate to the socially embedded concepts constructed by our system given some initial topics. To do so, we calculated the percentage ρ of people that guessed the initial topic given the set of automatically constructed definitions. We then calculated the mean value μ and the standard deviation σ for the votes obtained for each correct guess (quoted to 2 decimal places). The definitions provided can be found in the Appendix. We obtained the results shown in Table 5.1

We can hence conclude that the socially embedded concepts automatically obtained from a topic highly represent such a topic in an artistic sense.

The aim of sub-survey (ii) is to compare the rankings obtained on a given set of socially embedded concepts. The survey ranking was calculated by ordering the topics' associations one to ten, according to the mean rank from the participants. The system ranking was calculated according to the *relevance* of each association to the initial topics. Such relevance was calculated in terms of the size of the cluster corresponding to each association and to the proportion that each feature in the definition of this cluster has with respect to the total number of features in that cluster, as follows:

$$Ranking(Def(c_i)) = \frac{|E(c_i)|}{|E(c_1 \cup ... \cup c_n)|}) * \prod_{f\ in\ Def(c_i)} \frac{V(f)}{|Def(c_i)|}$$

where, as in Section 5.1, c_j is a cluster defined as a set of features, $Def(c_i)$ is the set of features in the definition of cluster c_j, $E(c_j)$ is the set of of exemplars in cluster c_j and $V(f)$ is the number of times a feature f is encountered. The lists of socially embedded concepts for each topic used in the survey, and the corresponding rankings can be found in the Appendix. In order to establish the correlation between the two rankings, we calculated Pearson's correlation, r, between the system's ranking and an aggregated ranking. We then calculated the respective 95% Confidence Intervals (CI) and p-values, using the alternative hypothesis that the correlations are greater than zero. We obtained the results shown in Table 5.2 (quoted to 3 decimal places):

Topic	r	p	95% CI
Grandmother	-0.381	0.138	[-0.715, 1]
London Olympics	0.659	0.019	[0.168, 1]
War	0.345	0.164	[-0.256, 1]
Love	0.212	0.278	[-0.385, 1]

Table 5.2: Results of sub-survey (ii): Pearson's correlation, r, p-values and 95% Confidence Intervals (CI) obtained by comparing the system's ranking and an aggregated ranking.

We can then conclude that there is a high and significant correlation between the rankings for the topic 'London Olympics', a weak and non-significant correlation between the ranking for the topics 'War' and 'Love', and a negative and non-significant correlation for the topic 'Grandmother'.

5.3   Conclusions and Further Directions

In this chapter, we have studied a method for generating socially embedded concepts related to an initial topic. The algorithm applies the Rational Model of Classification [5], a concept formation method widely used in cognitive psychology, to a set of data obtained from Twitter [1]. We then analysed the obtained concepts by :

  (i). asking a set of 25 people to first guess a topic given four sets of socially embedded concepts automatically created, and then to rate how well these concepts represent the topic from an artistic point of view.

(ii). asking a set of 25 people people to rank four sets of ten socially embedded concepts, each automatically created from one of four initial topics.

The experiments reported in Section 5.2 indicate that our system is capable of generating socially embedded concepts that could be of artistic interest to an audience. This ideation process could hence be used at the heart of more sophisticated artefact generation systems.

The results from survey (i) show that people not only can easily relate to the pre-defined initial topic given a set of socially embedded concepts automatically derived from it, but also that such socially embedded concepts are regarded to be strongly artistically representative of such an initial topic. The results from survey (ii) show that on average there is a weak positive correlation between the rankings obtained from our system (as described in Section5.2) and the ones provided by a sample of people.

We can hence conclude that the proposed method constitutes a good start towards the creation of algorithms for the automated constructions of socially embedded concepts. The results from survey (i) suggest that the concepts obtained from applying our algorithm to four initial topics could be used for the automated creation of artefacts, as they are considered to strongly represent the initial topics.

However, the weak correlations obtained from the analysis of the results from survey (ii) suggest that the rankings obtained from our system are not representative of public opinions. We hence believe that the system should be expanded by including automatic measures of social relevance. These could be based on both the measure of interestingness proposed in Chapter 4 and the notion on interestingness and curiosity reviewed in Section 2.5. For

example, the notion of *novelty* and the related belief that interesting concepts are 'similar-yet-different' to those that have been experienced previously [68], could be applied to socially embedded concepts. In doing so, concepts related to a topic would be considered interesting when it is 'similar-yet-different' from the average associations that the public uses for this topic. Moreover, similarly to our fictional concepts, we believe that the notion of *stimulation* should be captured by the evaluation. Stimulation in this case could be related to the likelihood that a tweet is 're-tweeted' or mentioned in another tweet.

Chapter 6

Conclusions and Contributions

In this thesis, we explored the field of automated artefact generation in computational creativity with the aim of proposing and analysing some methods of creation and evaluation of ideas of cultural value. We define an idea as being a concept which can be used to guide the generation of an artefact. In particular, we focused on two different kinds of ideas: fictional concepts and socially embedded concepts. In Chapters 4 and 5 we studies the ideas obtained by running two different methods of concept generation, one for each of the two kinds of ideas taken under consideration. We compared our results with the outcomes obtained from two sets of surveys. Both of our methods of idea generation make use of the notion of typicality, widely used in concept formation theories from cognitive psychology. Typicality is a measurement on the extent of belongingness of an exemplar to a concept. We believe that the use of typicality is highly relevant in computational creativity as it has been demonstrated that this factor is central for any flexible and subjective concept formation theory.

For the generation of fictional concepts, we used a well established theory formation system, HR [20]. One of the features of HR is the generation of non-existence conjectures. These are logical statements that are not satisfied by any entries in a given knowledge base. Such non-existence conjectures were used as a base for the definition of our fictional concepts. We then used typicality to assign to each of the newly generated fictional concepts three different dimensions of interestingness: novelty, vagueness and stimulations. The results obtained from the comparison between the system and a participants' rankings (with respect to these three measures of interestingness) show that both our measurements of novelty and vagueness respect the public beliefs. On the other hand, some improvements are still necessary for a valid measurement of stimulation.

For the generation of socially embedded concepts, we applied a typicality-based classification method, the Rational Model of Classification (RMC), to a set of data obtained from Twitter. In this case the scope was to create a set of concepts that naturally associate to an initial topic. The RMC was applied to four sets of tweets, downloaded using an external tool. Each set of tweets corresponded to one of four initial topics: 'Grandmother', 'London Olympics', 'Love' and 'War'. The result was a set of clusters per each topic, each cluster having a definition consisting of a set of words that appeared recurrently in the tweets. These define socially embedded concepts related to the initial topics. A survey asked people to first guess the topic given a set of definitions, and then to rate the artistic relevance of these definitions. The results showed both high association percentage and high relevance scores. A second survey was used to compare the rankings on the social impact of each of the definitions. The system rankings were based on the relevance of the cluster to the initial topic. The results obtained show a weak positive

correlation between the two rankings.

Our experiments show that it is possible to automatically generate ideas with the purpose of using them for artefact generation. This is an important step for the automation of computational creativity, because, to our knowledge, most of the available artefact generation systems (such as those reviewed in Section 2.2) are based on ideas either introduced by the programmer or directly extracted from a piece of humanly produced text. Moreover, our experiments introduce new ways of using the notion of typicality in computational creativity and show how these uses can lead to positive results for both the generation and evaluation of an ideas. However, whereas our results show a promising start, a lot of improvements and additions need to be considered. We analyse below the key findings from our studies, with the aim of building an initial framework that further research can refer to.

6.1 Key Findings and Further Directions

Below we report some findings arising from the studies reported in Chapter 4 and in Chapter 5. In Section 6.1.1 we draw some conclusions and theorise a framework for the ideation process. In Section 6.1.2 we draw some conclusions on the use of the notion of typicality in computational creativity.

6.1.1 A Framework for Ideation

The observations derived from the above chapters lead to a programme for ideation that has at its heart a series of questions. The rest of this section will discuss this programme. The material in this section is interlinked to

the proposal of the WHIM project, a three year European Initiative funded as a STREP (short-term research project) by the FP7 programme of the European Commission [3], to which we have contributed.

Firstly, we would like to expand once again on the key proposition behind this thesis. As reported in Chapter 2, in the field of Computational Creativity, the research that focuses on the implementation of autonomous creative systems usually addresses the methods of generation of artefacts in particular domains. The majority of this research has been devoted to designing software able to produce finished artefacts, without the software explicitly undertaking idea generation. In these cases, people are naturally inclined to themselves read ideas embedded in the artefacts. For example; a poem generated using a template may contain enough information for a reader to interpret a novel idea about the world described in the poem, but it is in fact the reader who provides the creative idea here, not by the software [3]. In certain areas of Artificial Intelligence research, especially Machine Learning, concept formation is the point of the exercise, and such concepts are a type of idea. However, the concepts formed tend to be used to describe and categorise real-life data. Hence, they were not designed explicitly for the purpose of provoking thought in the same way that a painter or a writer might do it creating a painting or story [3].

In implementing the two ideation systems described in Chapters 4 and 5 with the above scope in mind, we have noticed some common themes that we believe is worth underlining. These findings shouldn't be taken as absolute or final instructions, but as a framework on which further work could be based.

From the results of the two experiments reported in Chapters 4 and 5, it is obvious that the ideas generated from a system are highly dependent on the

initial knowledge base. Whether the knowledge base is constructed (as in Chapter 4) or retrieved (as in Chapter 5), the ideas generated from it mirror, and hence are limited to, the facts it contains. It follows that the creation of the knowledge base should be regarded as a key step in the idea generation process. For example, imagine we would like to generate the fictional idea of 'birds screaming because they are scared of heights'. This would be possible by using our fictional idea generation method, but the initial knowledge base would need to report information about the fact that birds are often found in high places; that birds are animals; that humans are animals; that humans sometimes get scared of heights; that humans sometimes scream when they are scared; and so on. Similarly, if we would like to create a socially embedded idea about the Olympics in Beijing, this could be done using our method, but we would need to be able to retrieve a complete knowledge base of tweets that were posted while the Beijing Olympics were on. Hence, we believe that one of the initial questions that an ideation process should address is:

How can a knowledge base be constructed that will contain enough information to support ideation? [3].

Assuming that the knowledge base has been created, the key step for both of our ideation methods was to construct an algorithm able to constructively extract and manipulate facts from this knowledge base in order to generate ideas of the required type. Given the number of unsound or irrelevant concepts that can be generated from a given knowledge base, it is important that the methods focus on the generation of concepts which are coherent, sound and potentially interesting. This gives rise to another question which

94

an ideation method should focus on:

> What methods can be implemented which will reliably produce
> coherent, sound and potentially interesting ideas? [3].

Finally, from the experiments in Chapters 4 and 5, we have noticed the need to decompose the ideation process into two independent parts: *idea generation* and *idea evaluation*. In our case, in the idea generation phase, a large set of ideas of a particular kind is formulated. In the idea evaluation phase, the most interesting ideas out of this large set are selected, through the use of some measurements of interestingness. The idea evaluation phase is key to the process. As our negative results in Section 5.2 demonstrate, it cannot be skipped; in Chapter 5 we have used part of the outcomes derived from the idea generation method in order to evaluate the ideas, without separating the two phases (i.e. we used the size of the cluster to rank ideas, which was calculated during idea generation). This led to negative outcomes, as explained in Section 5.2. This leads us to the final question that an ideation process should address:

> How can software reliably estimate the potential interestingness
> of an idea in a particular context? [3].

We hence conclude that an idea generation method should ultimately be composed by the components, one per each of the proposed questions above: *formalisation*, *implementation* and *evaluation*.

The formalisation component should focus on the creation of an initial knowledge base, by addressing questions such as: which information should the

knowledge base contain and how will they be represented (i.e. first order logic, raw text, n-grams etc.) The implementation component should be the heart of the system and should focus on the generation of ideas of a pre-determined type. It should address questions such as: what kind of ideas do the software aim to generate? How can the software generate such ideas so that they are sound, coherent and potentially interesting? The evaluation component should focus on methods to estimate the artistic value of a set of generated ideas. It should address questions such as: what are the key notions that would make an idea of a particular kind interesting the public? How can these notions be formalised into an evaluation formula? The three components don't need to be utilised in a linear fashion (i.e. be used one after the other, as we have done in our methods). They could instead pass information one to the other, in order to improve their partial processing. For example, the implementation component may go back to the formalisa-tion component in order to ask more world view knowledge information, or the evaluation component could feed back to the implementation component in order to guide the generation process toward an area of high interest.

We believe that a final step should be added to the framework. This is nec-essary in order to bind an ideation system to its ultimate scope - to entertain people. We hence believe that ultimately ideas needs to the presented to and evaluated by the public. We call this final step the *Audience Embracement*. In the methods proposed in Chapter 4 and 5 the audience embracement phase coincides with the conduction and analysis of surveys. However, many other human oriented methods are possible. The human embracement step can be used not only to estimate the impact of ideas on the public, but also to study the affects or nature of human reactions. Its outcome can then be re-utilised by any of the three components of ideation. For example, if the human em-

bracement phase reports that the general reaction to an idea is confusion, this can be fed back to either the generation component, to lead the idea formation process through less complicated steps, or to the evaluation component, to re-tune the evaluation of ideas with respect to confusion, and so on. We have summarised the proposed components and their interactions in Figure 6.1.
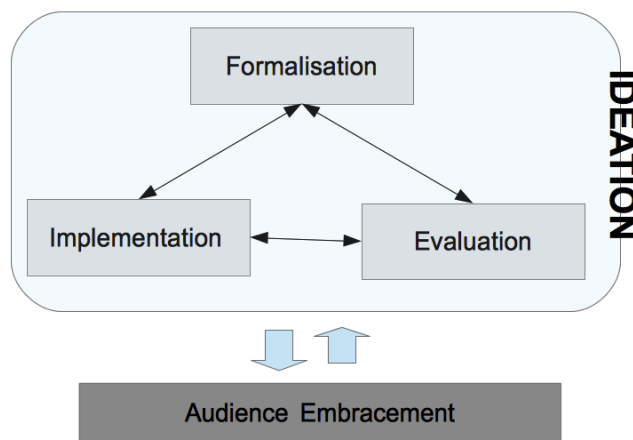
Figure 6.1: The proposed components of the ideation process.

By looking closely at the three ideation components proposed above, we can notice a parallelism with the commonly used four *P's of creativity* (reviewed in Section 2.4.2). The formalisation component can be associated with the creative Person, where the knowledge base can be thought of as the memory of an individual. The implementation component can be trivially associated with the creative Process; both focus on the creation of ideas. Finally, the evaluation component can be associated with the creative Product, as both address the estimation of the value of a product/idea.

6.1.2 Typicality in Computational Creativity

Another contribution of this thesis to the field of Computational Creativity is the use of the concept of *typicality* for both the generation and evaluation of ideas. As a reminder, and as reviewed in Section 2.3, typicality is a notion widely used in concept formation theories in cognitive science. It is used to indicate the degree of membership of an exemplar to a category. For example, we could say that a Labrador is a more typical exemplar of the category 'Dog' than a Chihuahua, or that a stallion is a more typical exemplar of the category 'Horse' than a rocking horse. Typicality is usually presented as a mapping between an item and a value $\in [0, 1]$. Typicality is used in most theories of categorisation, and it is considered to be a key aspect of cognitive psychology not only because of its tangible features, but also because it is a defining aspect of the difference between people categorisations. Typicality is in fact believed to be interlinked with the definition of a category itself, where every time we classify a new item as a member of a category, the definition of this category changes. Typicality and category definitions are therefore dependent on an individual's memory and experience. For example, say that in a particular person's memory there is a category for 'Dog', and that this category's representation implies that dogs have four legs. If this person then sees a dog which has lost a leg, the representation of this category would change itself, and this person would be more inclined to classify objects with three legs as dogs.

The idea of using degrees of membership has been applied in Computational Creativity before. In particular, we refer to the use of fuzzy sets in [63]. In this work, Ritchie revisits Wiggins' formalization [81] of Boden's ways of creativity [12] (both reviewed in Section 2.4.1). In [81], Wiggins defines U

as the set of all possible things (called Universe). He uses this definition: 'The universe, U, is a multidimensional space, whose dimensions are capable of representing anything, and all possible distinct concepts correspond with distinct points in U' [81]. A conceptual space C is a subset of U, and it is defined by two functions N and Q [1]. As explained in Section 2.4.1, generally speaking, N defines the set of items that can be considered to be instances of C, and it is a function that maps an item to a value $\in [0, 1]$. Instead, Q defines the ordered search according to which these instances are explored. If we define the set of all possible instances of a conceptual space C as E, then an additional function V is used to measure the 'value' (in our case interestingness) of the members of E. Wiggins initially associates a threshold value equal to 0.5 to N, by implying that given an item $i \in U$, this item is considered to be part to a conceptual space C if this conceptual space's membership function N applied to i returns a value bigger or equal to 0.5. In [63], Ritchie picks up on this restriction and underlines that an arbitrary threshold value α can be used instead of the fixed value 0.5. Hence, Ritchie re-defines conceptual spaces as fuzzy sets, whose membership function is defined by N.

In Chapter 4, our methodology consists of a simple application of this model: here each fictional concept can be thought of as a conceptual space, and each real world animal is assigned to each of these conceptual spaces with a degree of membership. Subsequently, in Section 4.2, we demonstrate that this membership value can be used not only for the definition of the conceptual spaces, but also for its evaluation (referring back to [63], we are applying Wiggins' theory at a meta-level, as we are establishing interestingness based on the

---

[1] Here we have adopted the notation used in [63] for consistency and clarity within the thesis.

typicality of a conceptual space and not of basic items as in [62]). Hence, we imply that the two functions N and V above are intrinsically linked. In saying so, we do not make the assumption that this is a demonstrated fact, but we believe that the exploration and formalisation of such a relationship could constitute a valuable subject for further studies.

In Chapter 5, the concept of membership of an item to a conceptual space is used in a different direction: in this case, the difference between fuzzy sets and typicality is exploited. Such a difference arises when the definition of a conceptual space is itself dependent on its items' membership values (the set of mappings $E \rightarrow [0,1]$ defined by N). In cognitive science, and for our specific case in the RMC (reviewed in Section 2.3.2), a membership function N, and hence the corresponding conceptual space C itself, continuously change as more items are assigned to C. This implies that conceptual spaces are themselves dependent on the order in which items are explored, and hence on Q above. In cognitive science, this is believed to be one the reasons behind the differences in people's categorisations, as it implies that memory plays a major part in the definition of category. We hence believe that a potentially interesting link could be defined also between N and Q. Once again, we are not implying the existence or the nature of such a relationship, but we believe that it is an interesting point for further studies, especially given it's relationship with personification.

We provide an example to further explain the two points raised above. Imagine that we have a painting a and we make a big hole in it, transforming it to a different item b. Also imagine we have a canvas with a hole, and we call this item c. Then, b would belong to the conceptual map of paintings with a degree of membership t_b. Our first observation suggests that t_b can

be used to define the interestingness of b. Now imagine that two different people see c. The first person sees c after having seen b, while the second person sees c without having seen b. Then the degree of membership of c to the conceptual space of paintings would be different for the two people. This would imply that the conceptual space of paintings itself is different for the two people, which implies that the interestingness of c is different for the two people. This is what our second observation suggests.

Note that in the observations raised above, we do not specify how these eventual additions to Wiggins' formalisation would modify the definition of exploratory and transformational creativity. We leave this as an open question for further research.

6.2   Summary

In Sections 6.1.1 and 6.1.2 we have discussed the results from Chapters 4 and 5 with the scope of integrating our research into the bigger picture of Computational Creativity. We can conclude that the material discussed in this thesis contributes to the field in two distinct ways:

- The proposition of methods for idea generation, and the integration of these methods into an ideation framework. Such a framework would consist of three components: *formalisation, implementation* and *evaluation.*

- The further integration of typicality into automated creative systems. We did this by demonstrating how typicality can be used for both the

generation and evaluation of ideas and by showing how such uses fit
into the formalisation of Boden's 'ways of creativity'.

Bibliography

[1] https://twitter.com/.

[2] http://twitter4j.org/en/index.html.

[3] http://www.whim-project.eu/proposal.html.

[4] http://www.wordreference.com/.

[5] J.R. Anderson. *The adaptive character of thought.* Psychology Press, 1990.

[6] R. Baumgarten, M. Nika, J. Gow, and S. Colton. Towards the automatic invention of simple mixed reality games. In *Proceedings of the AISB'09 Symposium on AI and Games*, 2009.

[7] D.E. Berlyne. *Conflict, arousal, and curiosity.* McGraw-Hill Book Company, 1960.

[8] D.E. Berlyne. *Aesthetics and psychobiology.* Appleton-Century-Crofts, 1971.

[9] K. Binsted, H. Pain, and G. Ritchie. Children's evaluation of computer-generated punning riddles. *Pragmatics & Cognition*, 5(2):305–354, 1997.

[10] K. Binsted and G. Ritchie. An implemented model of punning riddles. In *Proceedings of the National Conference on Artificial Intelligence*, pages 633–633. John Wiley & Sons Ltd, 1994.

[11] M.A. Boden. The creative mind: Myths and mechanisms (second edition). *Routledge*, 2003.

[12] M.A. Boden. Creativity in a nutshell. *Think*, 5(15):83–96, 2007.

[13] S. Bringsjord and D.A. Ferrucci. Artificial intelligence and literary creativity: Inside the mind of brutus, a storytelling machine. In *Proccedings of the AAAI Fall Symposium on Narrative Intelligence. AAAI, Lawrence Erlbaum, Mahwah*. MIT Press, 1999.

[14] F. Cavallo, S. Colton, and A. Pease. Uncertainty modelling in automated concept formation. In *Proceedings of Automated Reasoning Workshop*, 2012.

[15] F. Cavallo, A. Pease, and S. Gow, J.and Colton. Using theory formation techniques for the invention of fictional concepts. In *Proceedings of the Fourth International Conference on Computational Creativity*, page 176, 2013.

[16] W. Chamberlain and T. Etter. *The Policeman's Beard is Half-Constructed: Computer Prose and Poetry.* Warner Software/Warner Books, 1984.

[17] S. Colton. The Painting Fool. `http://www.thepaintingfool.com`.

[18] S. Colton. Experiments in meta-theory formation. In *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science*, pages 100–109, 2001.

[19] S. Colton. Automated puzzle generation. In *Proceedings of the AISB'02 Symposium on AI and Creativity in the Arts and Science*, 2002.

[20] S. Colton. Automated theory formation in pure mathematics. *Springer Verlag*, 2002.

[21] S. Colton. The Painting Fool: Stories from building an automated painter. *Computers and Creativity*, 2012.

[22] S. Colton, A. Bundy, and T. Walsh. On the notion of interestingness in automated mathematical discovery. *Int. J. Hum.-Comput. Stud.*, 53(3):351–375, 2000.

[23] S. Colton, J. Goodwin, and T. Veale. Full-face poetry generation. In *Proceedings of the 3rd International Conference on Computational Creativity*, 2012.

[24] M. Cook and S. Colton. Automated collage generation-with more intent. In *Proceedings of the 2nd International Conference on Computational Creativity*, 2011.

[25] M. Cook, S. Colton, and J. Gow. Initial results from co-operative co-evolution for automated platformer design. In *Proceedings of the 2012t European conference on Applications of Evolutionary Computation*, pages 194–203. Springer-Verlag, 2012.

[26] M. Cook, S. Colton, and A. Pease. Aesthetic considerations for automated platformer design. In *Proceedings of the 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2012.

[27] D. Cope. *Computer Models of Musical Creativity*. The MIT Press, 2005.

[28] M. Csikszentmihalyi. 16 implications of a systems perspective for the study of creativity. *Handbook of creativity*, page 313, 1999.

[29] L. De Raedt and K. Kersting. Probabilistic inductive logic programming. pages 1–27, 2008.

[30] G. Diaz-Jerez. Composing with melomics: Delving into the computational world for musical inspiration. *Leonardo Music Journal*, pages 13–14, 2011.

[31] B. Falkenhainer, K.D. Forbus, and D. Gentner. The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, 41(1):1–63, 1989.

[32] G. Fauconnier and M. Turner. Conceptual integration networks. *Cognitive science*, 22(2):133–187, 1998.

[33] G. Fauconnier and M. Turner. The way we think: Conceptual blending and the mind's hidden complexities. *Basic Books*, 2002.

[34] D. Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive science*, 7(2):155–170, 1983.

[35] P. Gervás. An expert system for the composition of formal spanish poetry. *Knowledge-Based Systems*, 14(3):181–188, 2001.

[36] J.E. Grady, T. Oakley, and S. Coulson. Blending and metaphor. *Community Mental Health Journal*, pages 101–124, 1999.

[37] J.P. Guilford. The structure of intellect. *Psychological Bulletin*, 53(4):267, 1956.

[38] J.C. Kaufman and R.A. Beghetto. Beyond big and little: The four c model of creativity. *Review of General Psychology; Review of General Psychology*, 13(1):1, 2009.

[39] A. Kilgarriff. Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2):135–155, 1997.

[40] George J. Klir. Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? *Fuzzy Sets and Systems*, 24(2):141 – 160, 1987.

[41] A. Kozbelt, R. A. Berghetto, and M. A. Runco. Theories of creativity. In J.C. Kaufman and R.J. Sternberg, editors, *The Cambridge handbook of creativity*, pages 20–47. Cambridge Univ Press, 2010.

[42] J.K. Kruschke. ALCOVE: an exemplar-based connectionist model of category learning. *Psychological review*, 99(1):22, 1992.

[43] A. Krzeczkowska, J. El-Hage, S. Colton, and S. Clark. Automated collage generation-with intent. In *Proceedings of the 1st International Conference on Computational Creativity*, 2010.

[44] D.B. Lenat. Automated theory formation in mathematics. In *Proceedings of the 5th international joint conference on Artificial intelligence-Volume 2*, pages 833–842. Morgan Kaufmann Publishers Inc., 1977.

[45] D.B. Lenat. Learning program helps win national fleet wargame tournament. *ACM SIGART Bulletin*, (79):16–17, 1982.

[46] D.B. Lenat. Eurisko: a program that learns new heuristics and domain concepts: the nature of heuristics iii: program design and results. *Artificial Intelligence*, 21(1):61–98, 1983.

107

[47] D.B. Lenat and J.S. Brown. Why AM and EURISKO appear to work. *Artificial Intelligence*, 23(3):269–294, 1984.

[48] D.B. Lenat, W.R. Sutherland, and J. Gibbons. Heuristic search for new microcircuit structures: an application of artificial intelligence. *AI Magazine*, 3(3):17, 1982.

[49] B.C. Love and D.L. Medin. Sustain: A model of human category learning. In *Proceedings of the National Conference on Artificial Intelligence*, pages 671–676. John Wiley & Sons Ltd, 1998.

[50] P. McCorduck. *Aaron's code*. WH Freeman and Co., 1991.

[51] W. McCune. *The OTTER user's guide*. Technical Report ANL/90/9, Argonne National Laboratories, 1990.

[52] W. McCune. *A Davis-Putnam program and its application to finite first-order model search*. Technical Report ANL/MCS-TM-194, Argonne National Laboratories, 1994.

[53] J.R. Meehan. Tale-spin, an interactive program that writes stories. In *Proceedings of the Fifth International Conference on Artificial Intelligence*, 1977.

[54] R.G. Morris, S.H. Burton, P.M. Bodily, and D. Ventura. Soup over bean of pure joy: Culinary ruminations of an artificial chef. In *International Conference on Computational Creativity*, page 119, 2012.

[55] G.L. Murphy. *The big book of concepts*. MIT Press, 2004.

[56] Diarmuid O'Donoghue and Mark T. Keane. A creative analogy machine: Results and challenges. In *Proceedings of the Third International*

Conference on Computational Creativity, page 17–24, Dublin, Ireland, 2012.

[57] A. Pease. *A Computational Model of Lakatos-style Reasoning.* PhD thesis, School of Informatics, University of Edinburgh, 2007. Online at http://hdl.handle.net/1842/2113.

[58] F. Pereira and A. Cardoso. The horse-bird creature generation experiment. *AISB Journal*, 2003.

[59] F.C. Pereira. *Creativity and artificial intelligence: a conceptual blending approach.* Walter de Gruyter, 2007.

[60] M. Rhodes. An analysis of creativity. *The Phi Delta Kappan*, 42(7):305–310, 1961.

[61] R. Richards. Everyday creativity and the arts. *World Futures*, 63(7):500–525, 2007.

[62] G. Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99, 2007.

[63] G. Ritchie. A closer look at creativity as search. In *International Conference on Computational Creativity*, pages 41–48, 2012.

[64] G. Ritchie and F.K. Hanna. AM: A case study in AI methodology. *Artificial Intelligence*, 23(3):249–268, 1984.

[65] E. Rosch. Natural categories. *Cognitive Psychology*, 4(3):328 – 350, 1973.

[66] S.J. Russell, P. Norvig, J.F. Canny, J.M. Malik, and D.D. Edwards. *Artificial intelligence: a modern approach*, volume 2. Prentice hall Englewood Cliffs, NJ, 1995.

[67] A. N Sanborn, T. L Griffiths, and D. J. Navarro. A more rational model of categorization. In *Proceedings of the 28th annual conference of the cognitive science society*, pages 726–731, 2006.

[68] R. Saunders. *Curious Design Agents and Artificial Creativity: A Synthetic Approach to the Study of Creative Behaviour*. PhD thesis, Department of Architectural and Design Science, Faculty of Architecture, University of Sydney, 2002.

[69] J. Schmidhuber. What's interesting? *Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale*, 1997.

[70] A. Silberschatz and A. Tuzhilin. What makes patterns interesting in knowledge discovery systems. *Knowledge and Data Engineering, IEEE Transactions on*, 8(6):970–974, 1996.

[71] E. E. Smith and D. L Medin. *Categories and concepts*. Cambridge, MA: Harvard University Press, 1981.

[72] R.J. Sternberg. *Handbook of creativity*. Cambridge Univ Press, 1999.

[73] O. Stock and C. Strapparava. Hahacronym: Humorous agents for humorous acronyms. *Stock, Oliviero, Carlo Strapparava, and Anton Nijholt. Eds*, pages 125–135, 2002.

[74] A.M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

[75] S.R. Turner. *The creative process: A computer model of storytelling and creativity*. Lawrence Erlbaum, 1994.

[76] W. Vanpaemel, G. Storms, and B. Ons. A varying abstraction model for categorization. In *Proceedings of the Annual Conference of the Cognitive Science Society*, 2005.

[77] Tony Veale. Exploding the creativity myth: The computational foundations of linguistic creativity. 2012.

[78] T.B. Ward and Y. Kolomyts. Cognition and creativity. In J.C. Kaufman and R.J. Sternberg, editors, *The Cambridge handbook of creativity*, pages 93–112. Cambridge Univ Press, 2010.

[79] Thomas B. Ward. Cognition, creativity, and entrepreneurship. *Journal of Business Venturing*, 19(2):173 – 188, 2004.

[80] G. Weinberg, A. Raman, and T. Mallikarjuna. Interactive jamming with shimon: a social robotic musician. In *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pages 233–234. IEEE, 2009.

[81] G. A. Wiggins. A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, 19(7):449–458, 2006.

[82] Geraint A Wiggins, George Papadopoulos, Somnuk Phon-Amnuaisuk, and Andrew Tuson. Evolutionary methods for musical composition. *International Journal of Computing Anticipatory Systems*, 1999.

Appendix

Concept Definition	Software Ranking	Survey Global Ranking	Survey Mean Ranking
An animal that has a body-part with which it can both see and eat	1	1	4.88
A mammal with feathers	2	4	7.11
A dolphin that lives on grass	3	11	7.89
A bird with tentacles	4	3	6.89
A bird with a trunk	5	10	7.58
A pig which is a bug	6	2	5.85
A fish with a trunk	7	7	7.37
An animal that lives both under freshwater and in the arctic	8	8	7.52
A fox which is an amphibian	9	9	7.54
A cow with tentacles	10	12	8.43
A fish which is also an otter	11	6	7.14
A salmon with feathers	12	13	9.82
A bat which is also a zebra	13	5	7.12
A gecko with spines	14	14	9.88

Table 6.1: Fictional concepts sorted from highest scoring to lowest scoring with respect to the software ranking for measure M_V, compared with the survey values for vagueness.

Concept Definition	Software Ranking	Survey Global Ranking	Survey Mean Ranking
A mammal that lives in the ocean that can fly	1	1	3.93
A mammal that lives in the ocean with wings	2	3	6.18
A mammal with wings that can be ridden by humans	3	2	3.94
A bird that lives in a forest that can swim under water	4	4	6.81
An invertebrate with legs that can swim under water	5	5	7.39
A mammal with wings that can hunt	6	7	8.11
A mammal that lives under freshwater and with fins	7	13	9.36
A mammal that lives both under freshwater and under the ocean	8	14	9.5
A mammal with fins that can hunt	9	12	9.24
An animal that lives both under freshwater and in a forest and that has wings	10	6	8.09
An animal that lives both under freshwater and in a forest and that has a fur	11	8	8.13
A bird that lives under freshwater and that can swim underwater	12	9	8.35
A bug that lives in a forest and has claws	13	11	9.14
A mammal with a tail that can fly	14	10	8.36

Table 6.2: Fictional concepts sorted from the highest scoring to the lowest scoring with respect to the software ranking for measure M_N, compared with the survey values for novelty.

Concept Definition	Software Ranking	Survey Global Ranking	Survey Mean Ranking
A fish with lungs	1	13	9.98
An animal that has eyes with which it can defend itself	2	3	5.88
A fish that can walk	3	7	7.22
An arachnid which is a mammal	4	11	8.85
A tiger with wings	5	2	5.85
An animal that lives under the ocean and that humans can ride	6	5	6.22
A wolf that can fly	7	4	5.97
A horse that lives under freshwater	8	10	8.27
A predatory bird with fins	9	12	9.19
A chicken that lives in the arctic	10	14	10.27
A dolphin which is also an arachnid	11	8	7.33
A chicken which is also a shark	12	1	5.3
An animal that has a body-part with which it can both see and eat	13	9	8.02
An animal with trunk with which it can fly	14	6	6.68

Table 6.3: Fictional concepts sorted from the highest scoring to the lowest scoring with respect to the software ranking for measure M_S, compared with the survey values for stimulation.

s

Association For 'Grandmother'	Software Ranking	Survey Global Ranking	Survey Mean Ranking
Happy birthday Grandma, 100 years old today	1	6.6	10
When my grandmother died	2	6.4	8
My grandmother gives me all that I want	3	3.4	1
My grandmother recipes are the best	4	5.4	5
Grandma passed away. I will miss you	5	6.5	9
My grandmother is getting old	6	5.6	6
My grandmother is just great	7	5.2	4
I love visiting my grandmother house	8	4.8	2
My grandmother is a woman	9	5.0	3
My grandmother has Alzheimer	10	6.2	7

Table 6.4: Socially embedded concepts associated to the topic 'Grandmother' sorted from highest scoring to lowest scoring with respect to the software ranking for relevance, compared with the sub-survey (i) results.

Association For 'London Olympics'	Software Ranking	Survey Global Ranking	Survey Mean Ranking
We will beat them	1	5.2	4
Go England	2	2.4	1
I got tickets for the game	3	5.2	4=
I will win a medal	4	4.8	3
Can't wait for the England France game	5	5.2	6
Can't wait for the Olympics	6	5.9	9
I am ready for the London Olympics	7	4.4	2
London Olympics gets a bailout	8	5.8	8
London Olympics: great overseas holiday	9	6.2	10
London Olympics cost 3x original budget	10	5.6	7

Table 6.5: Socially embedded concepts associated to the topic 'London Olympics' sorted from highest scoring to lowest scoring with respect to the software ranking for relevance, compared with the sub-survey (i) results.

Association For 'Love'	Software Ranking	Survey Global Ranking	Survey Mean Ranking
I love you	1	6.0	6
Kissing the person I love	2	5.2	4
Life is good when you find love	3	5.6	5
I felt in love	4	4.2	2
It's true love	5	4.4	3
Love is finding someone righte	6	6.4	8
Love is the greatest thing	7	7.0	10
Happy birthday, I love you	8	2.8	1
Love hurts too much	9	6.2	7
The moment you fall in love	10	6.8	9

Table 6.6: Socially embedded concepts associated to the topic 'Love' sorted from highest scoring to lowest scoring with respect to the software ranking for relevance, compared with the sub-survey (i) results.

Association For 'War'	Software Ranking	Survey Global Ranking	Survey Mean Ranking
During World War One	1	5.2	5
Fight against war crime	2	5.0	4
Soldier marching to war	3	6.4	8
I warned you this would be war	4	3.4	1
The war killed thousand of people	5	7.0	9
Somewhat regret starting this war	6	3.8	3
Remembering the war	7	5.8	6
Declaring civil war	8	6.2	7
War and national pride	9	3.5	2
Civilian and children die during war	10	8.8	10

Table 6.7: Socially embedded concepts associated to the topic 'War' sorted from highest scoring to lowest scoring with respect to the software ranking for relevance, compared with the sub-survey (i) results.