Using Theory Formation Techniques for the Invention of Fictional Concepts

Flaminia Cavallo, Alison Pease, Jeremy Gow, Simon Colton
Computational Creativity Group
Department of Computing
Imperial College, London

ccg.doc.ic.ac.uk

Abstract

We introduce a novel method for the formation of fictional concepts based on the non-existence conjectures made by the HR automated theory formation system. We further introduce the notion of the typicality of an example with respect to a concept into HR, which leads to methods for ordering fictional concepts with respect to novelty, vagueness and stimulation. To test whether these measures are correlated with the way in which people similarly assess the value of fictional concepts, we ran an experiment to produce thousands of definitions of fictional animals. We then compared the software's evaluations of the non-fictional concepts with those obtained through a survey consulting sixty people. The results show that two of the three measures have a correlation with human notions. We report on the experiment, and we compare our system with the well established method of conceptual blending, which leads to a discussion of automated ideation in future Computational Creativity projects.

Introduction

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 a knowledge base. For instance, in machine learning, learned concepts are tested for predictive accuracy against a test set of real world examples. 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 a creativity purposes is still largely unexplored.

The importance of evaluating concepts independently of their truth value has been highlighted by some cognitive science research. 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. In (Saunders 2002), interestingness is evaluated through the use of the Wundt Curve (Berlyne 1960), a function that plots hedonistic values with respect to novelty. The maximum value of the Wundt curve is located in a region close to the y-axis, meaning, as Saunders points out, that the most interesting concepts are those that are "similar-yet-different" to the ones that have already been explored (Saunders 2002). The notions of actionability and unexpectedness were first introduced in (Silberschatz and Tuzhilin 1996) as measurements of subjective interestingness. Actionability evaluates the number of actions or thoughts that an agent could undertake as a consequence of a discovery. 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 (Klir 1987).

The importance of *generating* concepts which describe contexts outside of reality was underlined by Boden when she proposed her classification of creative activity. In particular, Boden identifies 'three ways of creativity' (Boden 2003): *combinational creativity*, *exploratory creativity* and *transformational creativity*. 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 (Fauconnier and Turner 2002) 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 (Pereira 2007).

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 (Colton 2002b) (reviewed below), and on cognitive science notions of concept representation. In particular, we explore how the notion of *typicality* can improve and extend HR's concept formation techniques. 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 classical view (Rosch

1973), 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, and the typicality of an exemplar with respect to a concept can be assessed.

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 our approach and the well established methods of conceptual blending. Finally, we draw some conclusions and discuss some future work.

Automated Theory Formation

Automated theory formation concerns the formation of interesting theories, starting with some initial knowledge then enriching it by performing inventive, inductive and deductive reasoning. For our purposes, we have employed the HR theory formation system, which has had some success inventing and investigating novel mathematical concepts, as described in (Colton and Muggleton 2006). HR performs concept formation and conjecture making by applying a concise set of production rules and empirical pattern matching techniques respectively. The production rules take as input the definition of one or two concepts and manipulates them in order to output the definition of the new concept. For example, the *compose* production rule can be used to merge the clauses of the definitions of two concepts into a new definition. It could, therefore, be given the concept of the number of divisors of an integer and the concept of even numbers and be used to invent the concept of integers with an even number of divisors. The *success set* – the collection of all the tuples of objects which satisfy the definition – of the new defined concept is then calculated. Once this is obtained, it is compared with all the previously generated success sets and used to formulate conjectures about the new concept. These conjectures take the form of equivalence conjectures (when two success sets match), implication conjectures (when one success set is a subset of another), or non-existence conjectures (when a success set is empty).

In domains where the user can supply some axioms, HR appeals to third party theorem provers and model generators to check whether a conjecture follows from the axioms or not. 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. Each item in the agenda represents a *theory formation step*, which is an instruction about 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 in the theory, and the most interesting concepts are developed first. Overall interestingness is calculated as a weighted sum (where the weights are provided by the user) of a set of measurements, described in (Colton 2002b) and (Colton, Bundy, and Walsh 2000). These were developed to evaluate non-fictional concepts, but some of them could be modified to evaluate fictional concepts for our system, and we plan to do this in future work. 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 (Baumgarten et al. 2009), puzzles (Colton 2002a), HR's own theories (Colton 2001) and visual art (Colton 2008).

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 paper 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.

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 definition. 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.

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 fictional concepts have not been included into the HR program to guide concept formation.It is, however, our ambition to turn these measurements into measures of interest for ordering HR's agenda.

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 discussed, vagueness is a measurement that has been widely studied in the context of fuzzy sets. Klir (1987) 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 (Klir 1987), 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 $M_V(C_1) > M_V(C_2)$.

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 (Saunders 2002). 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.

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 (Ward 2004). 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? 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.

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 (Pereira 2007), which is described in the next section. The concept map for a horse was taken from (Pereira and Cardoso 2003) 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 (Pereira 2007). Figure 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 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 in (Colton 2002b). 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 2, 3 and 4 of the appendix respectively.

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

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 1: Details from the knowledge base for animals.

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 2, 3 and 4 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 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/vagueness: r = 0.552, p = 0.020, 95% CI = [0.124, 1]

M_N/novelty: r = 0.697, p = 0.003, 95% CI = [0.350, 1]

M_S/stimulation: r = -0.029, p = 0.059, 95% 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

Figure 2: Word clouds: vagueness, novelty and stimulation.

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 word 'novel' or 'vague'. Moreover, these meanings are often very distant from the interpretation of the term 'stimulation' that we used in deriving the M_S measure. In figure 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 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, which will be studied in future work.

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 1 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. In future, we plan to use these likelihood scores 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.

CONCEPT: isanimal(A,horse), pw(A,wing)	
Non-fictional concept	**Likelihood**
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 1: Non-fiction concepts with success sets overlapping with atypical exemplars of the given concept, along with their actionability.

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 (Pereira 2007), which is a conceptual blending system implemented on the basis of the theory presented in (Fauconnier and Turner 2002). 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 (Pereira 2007).

Divago has been successfully tested in both visual and linguistic domains (Pereira 2007). 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 (Pereira 2007).

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 purposely 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 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 (as we plan to do in the future), 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, and we plan to explore this possibility.

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 (Pereira 2007, 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, which we are planning to implement soon. In order to do so, we will 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 will 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 we have found is correlated with vagueness.

Conclusions and Further Work

We have proposed a method for generating and evaluating fictional concepts, using the HR theory formation system enhanced with typicality values. With the experiments above, 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. Finally, we compared our approach to the one based on conceptual blending in the Divago system, which placed our work in context and highlighted comparisons which will inform future implementations.

Our system is still at the developmental stage. The experiment above, however, indicates that it 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. Our next steps will therefore be to refine our current approach and implement new measures to estimate the interestingness of fictional concepts. To start this process, we will take inspiration from the notions analysed in (Colton, Bundy, and Walsh 2000) and used in the HR system, and modify them as appropriate. We will also look at other measurements suggested and used in Computational Creativity literature, such as Ritchie's criteria (Ritchie 2007). These, for example, could be used to assess the novelty of a fictional concept with respect to other fictional concepts.

We will then refine our measurement of typicality. To do so we hope to take inspiration from the theories proposed in cognitive science on the evaluation of the prototype theory and the weighting of category features. Each feature will 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, we aim to introduce the notion of the *distortion of reality*. This measurement will serve to calculate how many real world constraints a fictional concept breaks. We will start by studying two methods for the calculation of values related to this. The first method is inspired from (Pease 2007) and will be 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. We will also implement further methods for reasoning with fictional concepts. These methods will 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. We also plan to 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. We hope that such studies will help usher in a new era of idea-centric approaches in Computational Creativity as we hand over the creative responsibility for ideation to our software and address high level issues such as imagination in software.

Acknowledgments

We would like to thank the anonymous reviewers for the comments and suggestions we received. This research was funded by EPSRC grant EP/J004049.

References

Baumgarten, R.; Nika, M.; Gow, J.; and Colton, S. 2009. Towards the automatic invention of simple mixed reality games. In *Proc. of the AISB'09 Symp. on AI and Games*.

Berlyne, D. 1960. *Conflict, arousal, and curiosity.* McGraw-Hill Book Company.

Boden, M. 2003. *The Creative Mind: Myths and Mechanisms (second edition)*. Routledge.

Colton, S., and Muggleton, S. 2006. Mathematical applications of Inductive Logic Programming. *Machine Learning* 64(1):25–64.

Colton, S.; Bundy, A.; and Walsh, T. 2000. On the notion of interestingness in automated mathematical discovery. *Int. Journal of Human-Computer Studies* 53(3):351–375.

Colton, S. 2001. Experiments in meta-theory formation. In *Proc. of the AISB'01 Symp. on AI and Creativity in Arts and Science*.

Colton, S. 2002a. Automated puzzle generation. In *Proc. of the AISB'02 Symp. on AI and Creativity in Arts and Science*.

Colton, S. 2002b. *Automated theory formation in pure mathematics*. Springer.

Colton, S. 2008. Automatic invention of fitness functions, with application to scene generation. In *Proceedings of the EvoMusArt Workshop*.

Fauconnier, G., and Turner, M. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.

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

Pease, A. 2007. *A Computational Model of Lakatos-style Reasoning*. Ph.D. Dissertation, School of Informatics, University of Edinburgh.

Pereira, F., and Cardoso, A. 2003. The horse-bird creature generation experiment. *AISB Journal* 1(3):257.

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

Ritchie, G. 2007. *Some Empirical Criteria for Attributing Creativity to a Computer Program*. *Minds and Machines*, Springer, 17:76–99.

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

Saunders, R. 2002. *Curious Design Agents and Artificial Creativity: A Synthetic Approach to the Study of Creative Behaviour*. Ph.D. Dissertation, Department of Architectural and Design Science, University of Sydney.

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

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

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 2: 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 3: 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 4: 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.