Automating Fictional Ideation using ConceptNet

Maria Teresa Llano and **Rose Hepworth** and **Simon Colton** and **John Charnley** and **Jeremy Gow**[1]

Abstract. The invention of fictional ideas (ideation) is often a central process in producing artefacts such as poems, music and paintings in a creative way. Automated fictional ideation should, therefore, be of much interest in the study of Computational Creativity, but only a few approaches have been explored. We describe here the preliminary results of a new method for automated generation and evaluation of fictional ideas which uses *ConceptNet*, a semantic network. We evaluate the results obtained through a small study that involves participants scoring ideas via an online survey. We believe this approach constitutes a firm basis on which a more sophisticated model for automated creative ideation can be built.

1 Introduction

Ideation is a portmanteau word used to describe the process of generating a novel idea of value. Fictional ideation therefore describes the production of ideas which are not meant to represent or describe a current truth about the world. As such, they have many purposes, one of which is to possibly unearth new truths, and another of which is to serve as the basis for cultural creations like stories, advertisements, poems, paintings, games and other artefacts.

A major field of study within Computational Creativity research involves designing software that exhibits behaviours perceived as creative by (human) observers [4]. As an example, *The Painting Fool*[2] system [2] is an automated artist that has produced pieces which have been exhibited in real and online galleries. Similarly, we have developed a system that generates poems automatically [3], where the poem represents a response to articles from the Guardian newspaper. In both these cases, as in the majority of the systems developed so far within Computational Creativity research, there is no idea generation undertaken explicitly. In many projects, especially applications to natural language generation such as neologism production [14], which are communicative in nature, it is entirely possible to extract ideas from the artefacts produced. However, it is fair to say that the software used in these projects is not performing ideation in order to produce artefacts, they are rather producing artefacts which enable the reader to interpret them via new ideas.

In the creative arts and the creative industries, the production of fictional ideas around which to write stories, paint pictures or design advertisements, is an essential activity. Computers cannot yet come up with interesting ideas, and we would like to change that. The work presented here is part of the WHIM[3] project, where we aim to undertake the first large-scale study of how software can invent, evaluate and express fictional ideas.

Our first step was to determine what is meant by fictional ideas in the context of this project. An idea which makes sense as a fiction is not necessarily one which excites the mind, as we shall see with some of the ideas considered by participants in the study described below. For instance, the idea: *What if there was a chair with five legs?* is coherent and it has saliency and is largely fictional, given that most chairs have three or four legs only. However, it takes some work to imagine a scenario in which a five-legged chair would be of particular interest. Hence, this idea is unlikely to enthuse people to play around with it in their mind by dreaming up humorous or dangerous or ridiculous scenarios in which the idea features. Moreover, it is likely that people could devise few narratives featuring a five legged chair as a central concept. A good fictional idea distorts the world view around it in useful ways, and these distortions can be exploited to spark new ideas, to interrogate consequences and to tell stories.

To illustrate these points, the ideas below represent one line summaries of the plots of two well-known stories:

> *What if we could give life to a being created by combining the body parts of dead people?*
>
> *What if there are other worlds, running parallel to ours, which can only be accessed by children?*

We can describe such ideas as being rich in narrative potential. That is, they might provoke a number of scenarios that are narratively interesting and which excite the imagination. However, it is important to note that audience appreciation of the value of an idea is often relative to the way in which the idea is presented, and the context in which this presentation occurs. For example, the What-if formulations presented above are both the basis of successful science fiction/ fantasy narratives. In contrast, consider the idea:

> *What if a Professor of Phonetics makes a bet that he can take a working-class flower seller and transform her into a genteel woman who can pass for a duchess?*

This may prove a popular narrative in a different context (indeed, it has), but the literal idea would likely have relatively little narrative potential in the context of the science fiction or fantasy genres[4].

We describe here a relatively basic ideation system that uses the ConceptNet semantic network [8], which is described in section 2. Our approach involves extracting certain facts from ConceptNet, hand-crafting an inversion of the reality expressed in the fact and wrapping it in an evocative rendering, as described in section 3. Using the results from an application of this technique, we have made an initial investigation into how to automatically estimate the narrative potential of fictional ideas. This involves enumerating inference

[1] Computational Creativity Group, Department of Computing, Goldsmiths, University of London. ccg.doc.gold.ac.uk

[2] www.thepaintingfool.com

[3] www.whim-project.eu

[4] Scenarios of existent stories may resemblance (parts of) the original idea; however, we refer to the idea as a whole as having little narrative potential on the science fiction or fantasy genres.

chains from ConceptNet, and testing whether the number and length of inference chains is an indicator of higher narrative potential.

In investigating the ideation and evaluation approaches, we have examined which of the idea generation strategies available with this method might be considered best to produce results with more narrative potential. In addition, we are interested in how the concept of narrative potential can be used as a reliable and measurable assessment method for fictional ideas overall. To do this, we plan a future large-scale crowd-sourcing exercise where people are exposed to automatically generated ideas in a controlled way. We report here on a small preparatory study for this exercise, as discussed in section 4, wherein we surveyed 10 participants' responses to a series of What-if style ideas. From the results and discussions arising from this survey, we draw tentative conclusions in section 5 which we hope will be of value for the crowd sourcing exercise. We conclude by describing some future developments for automated fictional ideation.

2 Background

Automated techniques for the derivation of new concepts have been important for a variety of Artificial Intelligence techniques, most notably Machine Learning [10]. However, the projects employing such techniques have almost exclusively been applied to finding concepts which somehow characterise *reality*, rather than some fictional universe. While some of the concepts may be purported as factual, e.g., supported by sufficient evidence, others may only be hypothesised to be true. In either case, however, the point of the exercise is to learn more about the real world through analysis of real-world data, rather than invent fictions for cultural consumption.

One project where the automatic generation of fictional rather than factual concepts was the aim is described in [11]. Here, Pereira implemented a system based on the psychological theory of Conceptual Blending put forward by Fouconnier and Turner in [5]. By blending theories about different subject material, novel concepts which exist in neither domain emerge from the approach. A classic example of this is Pegasus, the winged horse character of many stories, which arises as a blend of a bird and a horse. Using blending to reason about such fictional ideas has been harnessed for various creative purposes, including natural language generation [13], sound design [9], and the invention of character models for video games [12].

We have developed an automated fictional ideation approach using ConceptNet,[5] a semantic network of commonsense knowledge produced by sophisticated web mining techniques at the MIT media lab [8]. Mined knowledge is represented as facts, which comprise relations between concepts that are expressed as words and short phrases, in a network-like structure. There are many relations, including:

Antonym, AtLocation, CapableOf, Causes, CreatedBy, Desires, HasA, HasProperty, IsA, InstanceOf, LocatedNear, MadeOf, MemberOf, NotHasA, NotIsa, PartOf, SimilarTo, Synonym, UsedFor

Each fact is given a score from 0.5 upwards, which estimates the likelihood of the relation being true based on the amount of evidence mined. We extracted the bare information from ConceptNet into a set of tuples of the form: [LHSConcept,Relation,RHSConcept,Score]. As examples, the following are facts in ConceptNet about particular animals: [camel, IsA, animal, 7.0], [bee, CapableOf, make_honey, 2.0], [cat, Desires, play_with_string, 6.0], etc. Some relations are included in many facts, while others are included in far fewer.

[5] `conceptnet5.media.mit.edu`

In [8], Liu and Singh describe the various uses for ConceptNet, including finding contexts around a concept, making analogies and constructing chains of inference. The latter of these is of interest here. Liu and Singh provide an example of such a chain:

> ConceptNet can generate all the temporal chains between "buy food" and "fall asleep". One chain may be: "buy food" → "have food" → "eat food" → "feel full" → "feel sleepy" → "fall asleep". Each of these chains can be seen as being akin to a "script." . . . By knowing that "buy steak" is a special case of "buy food", . . . we can now make the inference "fall asleep".

An inference chaining approach has been used in the Emotus Ponens system, described in [7], for affective text classification. As described below, we similarly employ such chains to estimate the narrative potential of fictional ideas. As an implementation infrastructure for this, we have used a flowcharting system described in [1]. Providing details of how this system works is beyond the scope of this paper, but, of course, we give the details of the individual flowchart nodes we have employed, in order to present our approach.

3 Using ConceptNet for Fictional Ideation

As mentioned earlier, a good fictional idea distorts the world view around it. Thus, the first step for automatic ideation was to study ways of achieving such distortions. After identifying some common fictional ideas within well known stories or written by people on Twitter, we concluded that a straightforward method which *inverts* aspects of reality would be a good place to start. We have identified some general schemas through which this can be accomplished:

1. Stopping an action or desire which was previously common, widespread, fundamental and/or important. For example, 'people need to eat' becomes: *What if people no longer needed to eat?*
2. Equalising a property amongst something previously variable. For example, 'not everyone is pretty' becomes: *What if everyone was pretty?*
3. Starting an action which was not previously possible. For example, 'people can't fly' becomes *What if people could fly?*

The approach we have developed applies this reasoning by transforming ConceptNet facts via a flowchart represented in figure 1 below:

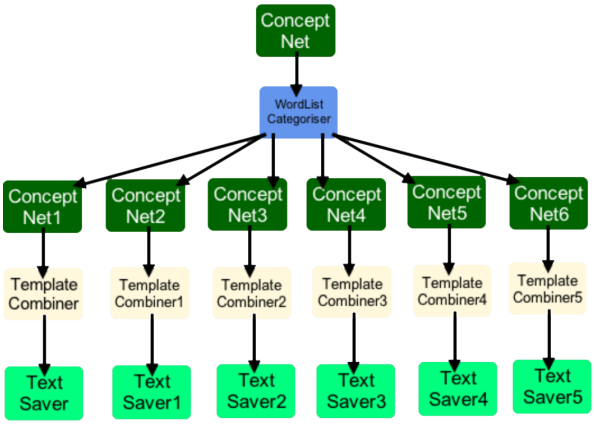

Figure 1. Flowchart for the ConceptNet-based fictional ideation process

More specifically, in order to generate fictional ideas in the form of *What-if* sentences, we alter the relations expressed by ConceptNet facts. The following steps are applied sequentially to achieve this:

1. The idea generation process is focused on a particular **theme** by finding all the terms X with a particular characteristic Y. In other words, ConceptNet is searched for facts [X,R,Y,S] – where R is a relation, X and Y are the left and right hand side of the relationship respectively, and S is the score associated to the fact, which is above a given threshold. For instance, all the concepts tagged as being an *animal* can be found in ConceptNet by identifying all the tuples of the form [X,IsA,*animal*,S], i.e., where R=*IsA* and Y=*Animal*. Choosing the score threshold involves trial an error based on the data retrieved by ConceptNet. For maximum yield, we have largely chosen to work with a threshold of 0.5. This step is carried out in the top node of the flowchart in figure 1.
2. After the theme selection, the next step is to **remove spurious data**. As ConceptNet facts are mined from the web, some inconsistent or incorrect data, such as "apple IsA animal" is sometimes found in its database. There are also facts which are true but not useful within the theme, e.g., "human IsA animal". In this step, we filter out such data by hand-crafting the parameters for the WordListCategoriser in the flowchart of figure 1, telling it to keep only the useful facts.
3. The next step in the process is to **select relations** relevant to the theme and use them produce ideas. That is, given a particular relation R_i, like *CapableOf* or *Desires*, find all the ConceptNet facts involving the previously selected terms X and one of the chosen relations. More specifically, find facts with the form [X,R_i,Z,S_i], where Z is a concept associated to X by R_i. Again, choosing the right score S_i is a trial and error process. These second appeals to ConceptNet happen in the row of 6 ConceptNet nodes in the flowchart of figure 1.
4. For each fact found in the previous step, this step involves **altering its reality** by transforming the relation R_i in the fact to form the What-if ideas. This is done by following an inversion scheme like those presented above. For instance, the fact [bee,CapableOf,make_honey] becomes *What if there was a bee who couldn't make honey?* – where schema 1 has been applied, i.e., stopping an action that was previously common.
5. In order to increase the potential value of the What-if ideas, further **rendering** of them is done by modifying parts of their statement. This can be achieved by assigning properties to the subject of the sentence. To illustrate this, adding the word *little* in front of an animal-centric idea yields ideas such as: *What if there was a little bee who couldn't make honey?* We believe this version of the idea would get a more emotional response, resulting in a better received idea. We hand-crafted templates able to do this for each of six relations chosen in step 3. Both the alteration of reality and the rendering is done in the TemplateCombiner nodes of the flowchart, followed by the saving of the resulting ideas to file.

3.1 A Disney Character Theme

Most facts about reality can have their truth inverted to produce a fictional idea. However, in order to test the value of such ideas, we needed a well-known context where such reality distortion is commonplace. One such context is the characters in children stories, and to further focus matters, we looked at anthropomorphised animals in Disney movies. Such characters are obviously fictional, and quite often there is an underdog theme involving an inability to perform a basic function which is fundamental to the general character of the animal type. The plot of the film often involves the character learning a particular skill, or succeeding without it to save the day/world/girl/boy/etc. As examples of the underdog meme, in Toy Story, Buzz Lightyear can't fly (even though he is a toy spaceman), Nemo the clownfish has trouble swimming, and the monsters in Monsters Inc. aren't particularly scary.

To produce ideas in this theme, we started with a knowledge base of animals obtained from ConceptNet and transformed facts about them through the procedure explained above. The results, broken down into the six ConceptNet relations we used, were as follows:

- *CapableOf* (116 ideas): negating abilities of animals, rendering each transformed fact as "What if there was a little X who couldn't Y?", e.g., *What if there was a little dolphin who couldn't swim?*
- *Desires* (83 ideas): negating what animals like to do, rendering each transformed fact as "What if there was a little X who was afraid of Y?", e.g., *What if there was a little cat who was afraid of drinking milk?*
- *LocatedNear* (39 ideas): negating common locations where animals tend to be, rendering each transformed fact as "What if there was a little X who couldn't find Y?", e.g., *What if there was a little ant who couldn't find the picnic?*
- *UsedFor* (91 ideas): negating what animals do, rendering each transformed fact as "What if there was a little X who forgot how to Y?", e.g., *What if there was a little bird who forgot how to nest?*
- *NotCapableOf* (55 ideas): negating what animals are not able to do, rendering each transformed fact as "What if there was a little X who learned how to Y?", e.g., *What if there was a little zebra who learned how to talk?*
- *HasA* (200 ideas): negating what animals possess, rendering each transformed fact as "What if there was a little X who lost its Y?", e.g., *What if there was a little dog who lost its tail?*

3.2 Automated Evaluation through Chaining

As per the individual figures for the number of ideas above, we generated a total of 584 What-if ideas describing Disney-like characters. To be of value as an ideation machine, software will need to automatically identify the most valuable ideas within such a set of candidates, and determining how best to do that will be an ongoing major challenge for the WHIM project. Part of the success of a fictional idea depends on whether the distortion of reality can be exploited to spark new ideas, to interrogate consequences and to tell stories. Given this, we developed a technique that automatically estimates the overall value of an idea by estimating its narrative potential.

The technique consists of building chains of relations whose starting point is the fact used to produce the idea. This kind of reasoning is possible in ConceptNet due to its graph-like structure, where all nodes are connected through relations, and transitivity can be used in order to form such chains. Based on this, we can evaluate an automatically generated idea by counting the number and lengths of possible chains of facts originating from it within the ConceptNet database. Each chain is considered as a possible narrative that could be developed from the original idea. To illustrate this, suppose we are given the original fact [bug,CapableOf,fly]. Then, from the seed idea *What if there was a little bug who couldn't fly?*, the following chain of relations can be obtained through ConceptNet:

[bug,CapableOf,fly]
↓
[fly,HasA,wing]
↓
[wing,IsA,arm]
↓
[arm,PartOf,person]
↓
[person,Desires,muscle]
↓
[muscle,UsedFor,move_and_jump]

One possible interpretation of this chain of facts is:

There is a little bug who can't fly, as he has arms instead of wings. He would develop arm muscles to move and jump instead of flying.

Through this interpretation, we could possibly imagine a Disney film about a little bug who, even though he cannot fly, overcomes adversity with super strength because of his muscular arms.

Automatically generating such interpretations is very much future work. However, such chains could still be of use. In particular, our hypothesis is that – while each chain might be rather poor and difficult to interpret as a narrative – the volume of such chains can indicate the potential of the idea. Hence our evaluation method gives ideas with more chains associated to them a higher score than those with fewer chains. To this end, we have developed a general strategy based on ConceptNet to construct chains of facts. This process consists of the following steps:

1. Form What-if ideas from ConceptNet facts [X,R,Y,S] following the procedure explained above – where the relation R is altered in order to form the What-if idea.
2. Choose a set of relations $\overline{R}$ to form possible parts of the chain.
3. Choose a minimum score *minS* to filter which facts will be accepted. This score serves the purpose of increasing the confidence that the chains generated can potentially be interpreted as narratives, and we normally choose a *minS* value of 1.0.
4. Choose a maximum size *maxSize* for the expected chains; i.e., the maximum number of relations a chain can contain. This effectively limits what could be a lengthy process, and we normally choose 12 as the limit.
5. For each fact [X,R,Y,S], search for facts that are connected through the right hand side of the relation, i.e., facts of the form [Y,R_2,Z,S_2] where:
   (a) $R_2 \in \overline{R}$; i.e., R_2 belongs to the set of allowed relations, and
   (b) $S_2 \geq$ *minS*; i.e., the score associated to the fact must be greater than or equal to the user defined minimum score.
6. For each of the retrieved facts, check that no cycles are formed if added to the chain, i.e., check that the pair (R_2,Z) – the relation and right hand side of the retrieved fact – does not already occur as the relation and left hand concept of a fact higher up the chain.
7. If no cycles are found, a chain is formed with the shape [X,R,Y,R_2,Z].
8. If the size of the new chain does not exceed *maxSize*, this one is then given as an input fact and the procedure starts again from step 5; i.e., the search focuses now on facts of the form [Z,R_3,W,S_3], and this continues until the maximum chain length is reached.
9. When all the possible chains up to length *maxSize* have been calculated for each of the ConceptNet facts under consideration, assign to it a score which is calculated to be the sum of the lengths of the chains starting from the fact.

In this fashion, we scored the 116 What-if ideas generated through the *CapableOf* relation. The idea scoring the highest was *What if there was a little fly who couldn't fly?*, with a score of 3,278,710. Along with 92 others, one of the ideas scoring the least was *What if there was a little bee who couldn't make honey?* with a score of 3, meaning that no chains were possible. It is likely that more stories could be imagined from the impossibility of flying than from the impossibility of making honey. Hence, at least when comparing extremes, this example supports the hypothesis that the scores assigned through the chaining process can be used to estimate narrative potential. We consider this hypothesis in the experiments described below.

4 Experiments and Results

To evaluate our approach, we conducted a survey in which a series of What-if ideas were ranked by participants with respect to certain qualities, with the rankings being translated into a score for each idea. The ideas produced were within the context of Disney films as described above, and we limited ourselves to using facts from ConceptNet with the CapableOf relation. We supplemented the ConceptNet ideas with a set of control ideas using a method where, before the inversion of reality and rendering stage, a ConceptNet fact had the right hand side replaced by a random verb. For instance, the fact $[dog, CapableOf, run]$ becomes $[dog, CapableOf, reckon]$ with the replacement of run with reckon from the verb 'to reckon'. We denote such random control ideas with R. From the ConceptNet produced ideas, we extracted some from those with No Chains *(NC)* and some from those with at least one ConceptNet Chain *(CC)*.

All the participants of the survey were native English speakers. This gave us confidence in the soundness of individual judgements by ensuring that the language used in the ideas presented was understood by all participants. This was sensible for this initial study, but we recognise that differences in interpretation through language and other factors will have to be taken into account in future.

4.1 Experimental Setup

Given the somewhat formulaic and well-known nature of the Disney character context, we were reasonably sure that the ideas would be understood well by all participants, and that the questions asked would be interpreted appropriately for the given context. The survey consisted of two parts, and there were two preparatory aspects in its development that tested both the success of ideation methods and the evaluative questions employed in the questionnaire. The first was to formulate questions that allowed us to gather consistently comparable responses enabling us to reliably measure the value people ascribed to the ideas as a cohort. The second was a question of data presentation: that is, selecting and submitting for evaluation the What-if ideas themselves.

Formulating survey questions appropriate for measuring narrative potential meant tailoring questions to the given context – in this case, characters central to an animated Disney film being pitched to a producer. In our first attempt at this, prior to conducting the study, we sought to break down the term *narrative potential* into constituent elements, the combined scoring of which would give an overall value for each idea. These questions were to be answered on a scale of 1 to 6, and were as follows:

- How fictional is this idea?

Table 1. Average participant scores for four questions, by class of idea: Random, Non-Chaining and ConceptNet Chaining.

Question	R	NC	CC
1.1. General impression	4.02	9.76	10.22
1.2. Emotional response	3.5	8.92	11.58
1.3. Level of surprise	10.62	6.82	6.56
2.1. Narrative potential	3.68	10.00	10.32

Table 2. Correlation between average general impression participant score and average participant scores for three questions.

Question	Correlation *(r)*
1.2. Emotional response	0.81
1.3. Level of surprise	-0.77
2.1. Narrative potential	0.87

Table 3. Correlation between ConceptNet fact score and average participant scores for four questions.

Question	Correlation *(r)*
1.1. General impression	0.18
1.2. Emotional response	0.15
1.3. Level of surprise	-0.61
2.1. Narrative potential	0.43

- How sophisticated is the language?
- How unusual is the phrasing of the idea?
- To what extent does this alter your perception of the animal?
- To what extent does this idea provoke an emotional response?
- How feasible is the scenario featured in this idea?

It became clear at an early stage that there were significant problems with these questions. Not least of these was that considering an individual score for a number of different character ideas in response to each of these six questions would be a laborious task, and we would struggle to manage fatigue. Furthermore, initial feedback to the questions indicated that many of them were too ambiguous to secure consistent interpretation. In particular, fictionality was interpreted either as being synonymous with feasibility, or it served as a short-hand for what we were calling narrative potential. Hence, in the latter interpretation, participants would be assessing all constituent elements of narrative potential in a single constituent question.

In rethinking the suitability of these questions, we decided that in asking people to rank the ideas from most successful to least successful (as an overall measure) in the given context, we were in fact asking them to rank them in terms of narrative potential. Rather than prescribing these in advance as above, we could then work out the constituent elements of narrative potential by asking respondents to rank the same ideas according to more specific questions. In doing so, we could measure the influence of these elements on respondents' general impressions by examining the degree of correlation between their general impression and their answers to subsequent questions.

In the first part of the questionnaire, we used the same set of 15 ideas presented as What-Ifs in three separate questions. The set of 15 were chosen by randomly taking five each from each of the R, NC and CC categories. For each of the three questions, the 15 were randomly shuffled in a different way. In the first question, we asked participants to rank the ideas in order of their general impression of each idea's overall success in regards to the given context (which was given as a preamble). We followed this by two further questions: we asked participants to rank the same list again according to (i) the degree of emotional response they felt upon reading and interpreting the idea, and (ii) the degree of surprise they felt in response to the idea. Feedback from the first formulation of the questions (as presented above) indicated that it was relatively easy to determine one's emotional response to these ideas, suggesting that this was a key component of respondents' general impressions of success. In asking people to consider the ideas in terms of the surprise they felt upon reading them, we were testing the hypothesis that feasibility and novelty would manifest themselves as a sense of surprise at the scenario in question. Assessing the degree of correlation between the first question and each subsequent question would, we reasoned, enable us to assess the component parts of narrative potential individually whilst calibrating them against the respondents' general impressions.

The second part of the survey was given to participants at least a day after the first part, to reduce any effect of fatigue. Here, we sought to investigate narrative potential directly, according to the number and quality of stories that each idea might generate in the imagination of the participants. In particular, we took two different lists of What-if ideas, and asked respondents to order each list according to the number and quality of the plot lines that they felt might be written about each of the featured Disney characters. The first list of ideas was identical to those used in the first part of the survey, although presented in a different order. This enabled us to check this question against general impression, whilst also comparing it to the results produced in response to the other questions in part one of the survey. The second list was constructed by sampling systematically at equal intervals in terms of chaining score across the set of ConceptNet-produced ideas which have at least one non-trivial chain. We recorded the score provided by the chaining method, so that a correlation between the score and participants' answers could be calculated. As usual, this list was randomly shuffled.

4.2 Results

Each of ten participants completed the survey. The average scores given for each class of ideas, i.e., Random (R), Non Chaining (NC) and ConceptNet Chaining (CC), are shown in table 1. Note that these averages correspond to the questions from the first part of the survey and question 1 only from the second part, and averages for individual questions are given in the appendix. Note that a rank of 1 (best) translated to a score of 15, while a rank of 15 (worst) translated to a score of 1, and the scores were averaged over all the participants.

These results show that, in general, for overall value, emotional content and potential for plot lines, the ConceptNet ideas were ranked as being significantly better than the random ones. Moreover, of the ConceptNet examples, those with chains scored slightly better than those without, but this might not be a statistically significant finding, given the low sample size. These results therefore add some support to our hypothesis that the ConceptNet chaining technique, that uses an estimate of narrative potential to rank ideas, provides a sound methodology to evaluate the potential of the ideas generated through our approach. Interestingly, in question 3 of survey 1, which assessed the level of surprise, the effect was reversed: the random ideas were ranked as best and the ideas with chains were ranked as worst in general. We believe this results from the interpretation of *surprising* by the participants of the survey when answering this question. We discuss this further in subsection 4.3 below.

Another aspect we evaluated was whether there was a correlation between general impression (question 1.1) and: (a) emotional response (question 1.2), (b) level of surprise (question 1.3) and (c) narrative potential (question 2.1). To this end, we calculated Pearson's product-moment correlation coefficient, *r*, between the average scores obtained from these questions. The results are shown in table 2. We see that there is a strong positive correlation between general impression and both emotional response and narrative potential. This confirms our hypothesis that both emotional response and narrative potential are key components of participants' general impressions of

value. However, there is a strong negative correlation between general impression and how surprising an idea was perceived to be. We believe this could be due to the interpretation of the question during the survey. As mentioned before, we expected that the concepts of feasibility and novelty would manifest as components of surprise; however, as formulated, the question may have been ambiguous and therefore specific questions about feasibility and novelty should be asked instead. We will take these findings (further discussed below) into account when designing the full crowd-sourcing exercise.

We also calculated the correlation between the average participants score for an idea and the score given by ConceptNet for the fact which was inverted for the idea. Ignoring the five randomly generated ideas, the correlations are given in table 3 for questions 1.1, 1.2, 1.3 and 2.1. From the final two correlations here, we can tentatively conclude that ideas appear less surprising when the ConceptNet fact about them is higher scoring, and that, when ideas from those with ConceptNet chains are presented, the narrative potential projected by people onto the idea will be somewhat in line with the score assigned by ConceptNet to the underlying fact.

Another specific objective of the study was to compare our chain scoring technique with the scores given by the participants. Question 2.2 was used for this purpose, where the 15 ideas in the list were automatically scored by the chaining approach, and chosen to at equal intervals in terms of the number of chains. To find the correlation between chain score and participants' average score, we also calculated Pearson's coefficient using the values of table 8 in the appendix. This resulting correlation is *r=0.23*. Although the correlation is weak, it is a positive, and considering that this is our first attempt to provide an automatic method for evaluating fictional ideas, we find this encouraging. We will explore the utility of the chaining scores for predicting the value of ideas as a central part of the full crowd-sourcing study.

4.3 Discussion

The first part of the survey evaluated the quality of the Disney characters portrayed by a set of What-if ideas. Regarding question 1, the idea of a little frog who couldn't jump was ranked best for general impression, while the idea of a little snake who couldn't tend was ranked at the bottom. This is consistent with our hypothesis about the value of the chain scores, since the first idea has ConceptNet chains, while the last was generated as a random idea. It seems that the ambiguity of the verb *tend* as well as the lack of context for the idea contributed to its poor result. More specifically, the verb *tend* means either *'having a tendency towards something'* or *'having to take care of something'*; therefore, more context is required. For instance, *What if there was a little snake who couldn't tend a bar?* may have had a better reception. Moreover, we believe the random aspect of the idea also affected the participants' response. In particular, a snake who cannot tend is not inverting a well-known reality, it is rather layering a fiction on top of another fiction, which may have been confusing or uninspiring.

Another aspect of the results that attracted our attention was that although the idea of a little frog who couldn't jump was ranked as best overall, the idea of a little frog who couldn't swim was much lower, ranked as ninth best for the same question. Our hypothesis is that jumping is a more definitive ability than swimming, i.e., we tend to associate frogs with jumping more than swimming, which could be because very few other animals also jump (regularly). Regarding question 1.2, which evaluated emotional response, the idea of a little whale who could not breath was ranked at the top. This is likely to be because people tend to feel more emotional when an idea is related to such a fundamental aspect of life as breathing, especially if that aspect is in jeopardy. In this example, the idea of a dying whale caused a more emotive response than the idea of a frog that cannot jump. However, a dying whale is not an appropriate character for a Disney film, as reflected in question 1.3.

In that third question, which evaluated the degree of surprise people felt about finding such characters suggested for a Disney film, the idea of a little snake who couldn't tend was ranked at the top. We drew different hypotheses regarding this finding:

1. *The character is very different to most Disney characters*. Usually, the main characters of Disney films are loving, caring and kind. The perception people have of snakes is, in general, very different. Moreover, this What-if idea proposes a lack or a skill that is not associated with a snake, and is not important to its well-being. To test if this hypothesis is true, we could ask participants to rank a list of ideas that includes characters that are different to the usual Disney characters, for instance a shark that cannot stop eating people, against a set of ideas which match the typical view. We could then ask participants to rank them based on how different these characters are to usual Disney characters.
2. *The question was interpreted as 'bizarre' rather than 'unusual'*. We observed that all the randomly generated characters were placed at the top of the ranking for this question. Hence it seems that the participants interpreted surprise as something bizarre or weird, instead of something unusual, like a cat who cannot cry. The only exception in the top ranking was the character of the whale who could not breathe, but as mentioned above, a dying whale in a Disney film is inappropriate and therefore, can also be interpreted as bizarre. To test this hypothesis we could ask participants to rank characters from most bizarre to least bizarre, by adding a set of bizarre characters to the list of ideas – these bizarre characters could be created using properties that do not relate to either animals or people (given the anthropomorphisation inherent in Disney characters). For instance, a dog who couldn't bend its bones would be a suitably bizarre construct.
3. *The question was interpreted as how surprising is it to see it on this list*. This could be because the question didn't highlight the nature of the ideas; i.e., it was interpreted at a meta-level as how unusual it is in the context of the questionnaire instead of the scenario of Disney films. We could address this by asking participants to rank the characters in the ideas on a novelty scale according to their view of commonly found characters in Disney films.

In addition to addressing the points above in future experiments, we plan to take into account the findings of Wundt [15], who points out that the hedonistic value of an artefact increases with novelty in the first instance, but then decreases as the novelty further increases, as it becomes more difficult to place the artefact into a context. Our findings here indicate that this is likely to be true of fictional ideas, and we hope to determine an automatic process which estimates the hedonistic value of an idea by estimating its novelty, so that it can present mid-novelty – high-hedonism – ideas as the best.

The second part of the survey evaluated the narrative potential of the What-if ideas. The main observation taken from the result of this question is that, as expected, the random ideas were ranked at the bottom. While fictional, an idea should make sense in order to be well received. The random ideas generally fail in this respect, since they are often completely out of context, i.e., they do not alter aspects of reality, since they describe relations which are not originally true.

As noted above, question 2.2 of the survey highlights a positive correlation between the automated ranking and the ranking given by

the participants of the survey. The correlation of 0.23 is weak, which might suggest that we need to improve the chaining technique, and we plan to try out variants of the approach in the full crowd sourcing exercise. On closer inspection, we see that the idea of a little bird who couldn't sing was ranked at the top by participants, while our ranking placed this idea at number 6. Analysing the scores given by participants – which can be found in table 8 of the appendix – it seems that the more common the association expressed by the original relation of an idea, the higher the score they receive from participants. That is, the value of an idea seems to increase in line with how strong the original aspect of reality is that is being transformed. For instance, it's more common to think of birds singing than to think of cats crying. The ConceptNet score for the underlying fact of an idea might give us a way to estimate the strength of the association, but the correlation in table 3 between that score and general impression is weak, so we cannot be conclusive at this stage about the value of the ConceptNet scores in assessing ideas overall.

5 Conclusions and Future Work

We have made two main contributions with this work. Firstly, we have investigated a generic approach for the automated generation of fictional ideas using ConceptNet. Secondly we have experimented with a technique to automatically score fictional ideas which can estimate their narrative potential. We have implemented both techniques through our flowchart system and have produced a set of 584 What-if style ideas that suggest fictional characters for Disney films. We also report on the results of an online survey in which participants were asked to rank fictional ideas based on general impression, emotional response, level of surprise and narrative potential. From the survey, we concluded that there is a strong positive correlation between the general impression participants have of the success of an idea and the aspects of emotional response and narrative potential, and a strong negative correlation between general impression and surprise. We have discussed this latter phenomenon, and plan to take it into account in future experiments. Moreover, the survey identified a small positive correlation between the scores assigned by the chaining approach and the average scores given by the participants.

Currently, our evaluation technique is based on chains of inference afforded by ConceptNet. Based on the correlation we found between general impression and emotional response, we could possibly improve the predictive power of the technique by using affective text classification techniques. An affect-based approach like that described in [7], where the affect of a concept is assessed through a chaining process, could be used to classify fictional ideas into affect categories, and this information used to good effect. We could follow a similar approach to that in [7], employing emotions commonly associated to a ConceptNet relation to determine the ideas more likely to be associated with stronger levels of emotion than others.

Choosing the right settings to produce ideas using the flowchart system was laborious. For instance, finding the relations, thresholds and chaining settings to generate interesting ideas took a few hours. We are currently enabling the system to automatically modify existing flowcharts to find high-yield configurations with little user intervention, and to invent entirely new flowcharts to work with databases such as ConceptNet. We would also like to look at the work of Doug Hofstadter and his CopyCat system [6], which enables the generation of creative analogies by discovering opposite concepts which can be associated, as a possible source for fictional settings.

As mentioned in the introduction, how an idea is presented can add to its value. We plan to implement rendering methods that will take narratives for an idea and produce interpretations of them which add value. We will experiment with the number and nature of the scenarios presented and test the hypothesis that presenting a moderate amount of supporting information for an idea can motivate people to expand the idea, and thus begin to own and appreciate it more.

The process for generating the ideas given above is generic, and we tested it by generating superhero-like characters, other types of cartoon characters and settings for surrealist paintings, with differing levels of success. In essence, we have presented here the first version of a *What-if Machine*[6] for fictional ideation. While it is certainly quite a basic prototype, we plan to build on this start, and use the results here as part of a baseline test suite against which we will chart our success. We believe that automated fictional ideation could lead to very useful software for the creative industries and beyond, and we have taken the first steps towards that with the work presented here.

Acknowledgements

We are very grateful to Stephen Clark and Mark Granroth-Wilding for their help in designing the study, and the members of the Computational Creativity Group at Goldsmiths and of the WHIM consortium, for their feedback on this work. We also thank the participants for taking the time to complete the survey. This project has been supported through EC funding for the project WHIM 611560 supported by FP7, the ICT theme, and the Future Emerging Technologies FET program, and also by EPSRC grant EP/J004049

REFERENCES

[1] S Colton and J Charnley, 'Towards a flowcharting system for automated process invention', in *Proceedings of the 4th International Conference on Computational Creativity*, 2013.
[2] S Colton, *The Painting Fool: Stories from Building an Automated Painter*, chapter 1 of Computers and Creativity, Springer, 2012.
[3] S Colton, J Goodwin, and T Veale, 'Full-FACE poetry generation', in *Proceedings of the 3rd International Conference on Computational Creativity*, 2012.
[4] S Colton and G Wiggins, 'Computational Creativity: The final frontier?', in *Proceedings of the 20th European Conference on Artificial Intelligence*, 2012.
[5] G Fauconnier and M Turner, *The way we think: Conceptual blending and the mind's hidden complexities*, Basic Books, 2008.
[6] D Hofstadter and M Mitchell, 'The Copycat Project: A Model of Mental Fluidity and Analogy-making', chapter 5 of Fluid Concepts and Creative Analogies, Basic Books, Inc., 1995.
[7] H Liu, H Lieberman, and T Selker, 'A model of textual affect sensing using real-world knowledge', in *Proceedings of the International Conference on Intelligent User Interfaces*, 2003.
[8] H Liu and P Singh, 'Commonsense reasoning in and over natural language', in *Proceedings of the 8th Int. Conference on Knowledge-Based Intelligent Information and Engineering Systems*, 2004.
[9] J Martins, F Pereira, E Miranda, and A Cardoso, 'Enhancing sound design with conceptual blending of sound descriptors', in *Proceedings of the Computational Creativity Workshop at ECCBR*, 2004.
[10] T Mitchell, *Machine Learning*, McGraw Hill, 1997.
[11] F Pereira, *Creativity and AI: A Conceptual Blending Approach*, Mouton de Gruyter, 2007.
[12] F Pereira and A Cardoso, 'The horse-bird creature generation experiment', *AISB Journal*, **1(2)**, 2003.
[13] F Pereira and P Gervás, 'Natural language generation from concept blends', in *Proceedings of the AISB Symposium on AI and Creativity in Arts and Science*, 2003.
[14] T Veale, 'Tracking the lexical zeitgeist with Wordnet and Wikipedia', in *Proceedings of the 17th European Conference on Artificial Intelligence*, 2006.
[15] W Wundt, *Grundzüge der Physiologischen Psychologie*, Engelmann, 1874.

[6] Which is the aim of the WHIM project: `www.whim-project.eu`

A Survey Results

In all the tables below, CC denotes a ConceptNet Chained idea, NC denotes a Non-Chaining ConceptNet idea and R denotes a randomly generated idea. Each table is organised in descending order of the average score given by the ten participants in the study.

Table 4. Average scores for ideas presented in question 1.1 (participants asked about general impression).

Idea	Score	Type
What if there was a little frog who couldn't jump?	14.5	CC
What if there was a little bird who couldn't sing?	13.0	NC
What if there was a little fly who couldn't fly?	11.9	CC
What if there was a little cat who couldn't catch a mouse?	11.0	NC
What if there was a little dog who couldn't eat a bone?	10.6	NC
What if there was a little dog who couldn't run?	9.6	CC
What if there was a little whale who couldn't breathe?	7.9	CC
What if there was a little cat who couldn't cry?	7.3	NC
What if there was a little frog who couldn't swim?	7.2	CC
What if there was a little bird who couldn't fail?	7.0	R
What if there was a little dog who couldn't wear a sweater?	6.9	NC
What if there was a little bird who couldn't reckon?	3.8	R
What if there was a little fly who couldn't chuck?	3.4	R
What if there was a little cat who couldn't fancy?	3.2	R
What if there was a little snake who couldn't tend?	2.7	R

Table 5. Average scores for ideas presented in question 1.2 (participants asked about emotional response).

Idea	Score	Type
What if there was a little whale who couldn't breathe	14	CC
What if there was a little bird who couldn't sing?	12.7	NC
What if there was a little fly who couldn't fly?	12.1	CC
What if there was a little dog who couldn't run?	11.2	CC
What if there was a little frog who couldn't jump?	10.7	CC
What if there was a little frog who couldn't swim?	9.9	CC
What if there was a little cat who couldn't catch a mouse?	9.9	NC
What if there was a little dog who couldn't eat a bone?	8.6	NC
What if there was a little cat who couldn't cry?	8.3	NC
What if there was a little dog who couldn't wear a sweater?	5.1	NC
What if there was a little bird who couldn't fail?	4.8	R
What if there was a little bird who couldn't reckon?	4.2	R
What if there was a little cat who couldn't fancy?	3.4	R
What if there was a little fly who couldn't chuck?	2.9	R
What if there was a little snake who couldn't tend?	2.2	R

Table 6. Average scores for ideas presented in question 1.3 (participants asked to indicate their level of surprise).

Idea	Score	Type
What if there was a little snake who couldn't tend?	12.6	R
What if there was a little fly who couldn't chuck?	11.5	R
What if there was a little whale who couldn't breathe?	10.7	CC
What if there was a little bird who couldn't reckon?	9.8	R
What if there was a little cat who couldn't fancy?	9.8	R
What if there was a little bird who couldn't fail?	9.4	R
What if there was a little cat who couldn't cry?	8.7	NC
What if there was a little dog who couldn't wear a sweater?	8.5	NC
What if there was a little fly who couldn't fly?	7.2	CC
What if there was a little dog who couldn't eat a bone?	7.1	NC
What if there was a little bird who couldn't sing?	6.2	NC
What if there was a little dog who couldn't run?	5.3	CC
What if there was a little frog who couldn't jump?	5.0	CC
What if there was a little frog who couldn't swim?	4.6	CC
What if there was a little cat who couldn't catch a mouse?	3.6	NC

Table 7. Average scores for ideas presented in question 2.1 (participants asked to estimate the potential for plot lines).

Idea	Score	Type
What if there was a little bird who couldn't sing?	14.4	NC
What if there was a little frog who couldn't swim?	13.0	CC
What if there was a little cat who couldn't catch a mouse?	12.5	NC
What if there was a little frog who couldn't jump?	11.1	CC
What if there was a little fly who couldn't fly?	10.9	CC
What if there was a little dog who couldn't eat a bone?	9.5	NC
What if there was a little dog who couldn't run?	8.8	CC
What if there was a little whale who couldn't breathe?	7.8	CC
What if there was a little cat who couldn't cry?	7.7	NC
What if there was a little bird who couldn't fail?	6.8	R
What if there was a little dog who couldn't wear a sweater?	5.9	NC
What if there was a little bird who couldn't reckon?	3.6	R
What if there was a little cat who couldn't fancy?	3.2	R
What if there was a little snake who couldn't tend?	2.9	R
What if there was a little fly who couldn't chuck?	1.9	R

Table 8. Average scores for ideas presented in question 2.2 (participants asked to estimate the potential for plot lines). potential, compared with the ConceptNet Chaining Score translated into a ranking score.

Idea	Score	Chain
What if there was a little bird who couldn't sing?	13.1	10
What if there was a little frog who couldn't jump	11.7	9
What if there was a little dolphin who couldn't swim?	11.1	14
What if there was a little bee who couldn't sting?	11.0	4
What if there was a little cat who couldn't see in the dark?	10.1	1
What if there was a little whale who couldn't sing?	8.7	11
What if there was a little dog who couldn't go for a walk?	8.5	6
What if there was a little dog who couldn't swim?	8.4	13
What if there was a little dog who couldn't run?	7.6	12
What if there was a little cat who couldn't cry?	6.6	15
What if there was a little whale who couldn't breathe?	5.8	3
What if there was a little dog who couldn't shake a hand?	5.1	2
What if there was a little dog who couldn't eat?	4.9	8
What if there was a little whale who couldn't reproduce?	3.9	7
What if there was a little dog who couldn't want?	3.5	5