A Framework for Exploring the Evolutionary Roots of Creativity

Hrafn Th. Thórisson

Latrastrond 7, 170 Seltjarnarnes, Iceland
h.thorisson@mac.com

This paper describes a development framework aimed at quantifying creativity and a simulation apparatus that allows its mechanisms to be traced to its evolutionary roots. The theoretical foundation of the framework is described, along with methods for addressing fundamental questions regarding the origins and nature of creativity and its relationship to logic. The method employs virtual worlds created using cellular automata. The worlds are inhabited by creatures with a cognitive system designed to resemble natural insects such as bees or ants. Creativity is operationalized by measuring the number of distinct plans used by the creatures; diversity of the environment is defined by the number of components that make up the environment and the total number of possible interactions between them. Primary hypotheses to be addressed are that environmental complexities govern the evolution of mental creativity and that creativity is bound by innate rules of the environment.

1 Introduction

Creativity is one of the most important abilities of humans. Our creative skills and ability to express it distinguishes us in obvious ways from the rest of the animal kingdom. Through creativity the human race has invented complex technology enabling it to acquire almost all abilities evident in other animals, as well as to go beyond them (e.g. interstellar travel, infrared vision). Considering the importance of creativity to our existence, surprisingly few studies have been directly aimed at understanding the general underlying structure of creativity mechanisms. By exploring the roots of creative mechanisms I hope to be able to understand by which means it is best to construct a general creativity system which can be applied in various different intelligence systems.

To get around the inherent limitations of human-subject experiments I am building a simulation environment for exploring alternative theories of creativity and setting up an environment where higher-level psychological variables and cognitive functions, such as perception and planning, are under direct control of the experimenter.

The primary hypotheses of interest here are that the evolution of creativity mechanisms are directly related to the complexity/diversity of the environment, and that the underlying mechanisms of creative behavior are essentially the same in all creatures, despite the obvious human advantages mentioned above. Depending on (a) the number of individual components in the environment and their interactions, (b) the creatures' ability to perceive their environment, and (c) the creatures' set of operators which can by applied to modify their behavioral pattern (in particular, to make plans), an increase in environmental diversity will result in a larger set of distinct behavioral patterns (plans) and, if the world is random, there will be no persistent structure evident in the composition of plans. [1] A secondary hypothesis to be addressed states that creativity is bound by rules inherent in the environment.

The hypotheses are grounded in the conjecture that creativity is evolution's answer to perceptually apparent unpredictability. This is different from some hypothetical or "actual" unpredictability – the discussion here revolves around unpredictability from the standpoint of the creature. A simplified version of the conjecture can be stated as follows: Creatures inhabiting a simple, closed and static world would tend to evolve to become completely robotic due to predictability of the environment; if an environment is very simple the cognitive system of the creatures require little or no effort to evolve mechanisms for survival. In a complex world, where events in the environment are not evident entirely by observation of the current situation, evolution must provide organisms with a mechanism to predict in uncertain situations. With the increase of interacting components in the environment, more complex cognitive efforts are thus required to produce and assess reactions to the current situation.

The paper is organized as follows: The related research and the theoretical foundation of the background hypotheses is discussed, followed by a short introduction to the terminology used in this paper. Then we present an example simulation environment for testing these hypotheses. At last we propose mechanisms for quantifying plan diversity and environmental complexity that will serve as the basis for quantitative measures of creativity.

2 Related Research

Peter Carruthers [2] has discussed the pretend play evident in children, where they enact adult situations – such as using a banana as a phone. He proposes that the function of the extensive creative play of children, also evident in the behavior of other mammalians where the young engage in pretend-play such as hunting and fighting, is to train the young in imaginative thinking for use in adult activities. I propose an extension to this hypothesis: During the initial stages of the development of more advanced animals such as mammals, creativity is yet to be bound by intrinsic rules of the environment. The gradual adjustment of creative mechanisms and internal representations to environmental structures impose logical reasoning and the ability to understand. The process of understanding, as defined by Baas & Emmeche [3], is related to, or equal to, explanation. Understanding complex adaptive systems (such as the Earth's environment) requires the discovery of causal relationships between different levels of organization. These different levels of organization are components and

---

[1] The complexity of the environment and the organism's ability to perceive the environment are equally important aspects of the evolution process, since varying complexity of the environment will make no difference if a creature is unable to perceive the variation. Vision is believed to constitute a significant proportion of information reception in humans, and I have selected the visual mode for my simulated creature's only sensual organ.

emergent, multi-leveled hyperstructures [3] which constitute the environment in a hierarchical manner, e.g. from atoms to raindrops to rivers.

Recent research has shed light on patterns spontaneously emerging at different scales in the synaptic-firing of neurons in the human brain [4]. They describe these patterns in the following manner:

"*These songs* [synaptic firing patterns] *resemble spiking correlates of sequential behavior, like bird songs or spatial navigation, and have compressing dynamics, as if the circuit was replaying and modifying previously learned sequencestivity ... the neocortex can spontaneously generate precisely reverberating temporal patterns of activation, dynamic ensembles that could represent endogenous building blocks of cortical function.*" [4].

These patterns could indicate the application of internal rule-structures to neural clusters representing external phenomena, aimed at the discovery of new conceptual relationships. The present framework provides a future foundation for observing and comparing any behavioral patterns which emerge through the composition of primitive actions.

Stephen Thaler [5] has created what he calls a "creativity machine" by perturbing connections in artificial neural networks and thereby creating 'noise' which manipulates learned concepts. A system inhabiting an ability to recognize abstract rules in the environment and use efficiently to manipulate conceptual structures (neural clusters) would behave similarly except for applying 'noise' in the form of "rule hyperstructures" to produce more ideas which correlate with the environment.

3 Theoretical Foundation

Earth's environment is governed by rules that can be traced down to a sub-atomic level, resulting in its emerging attributes such as land and sea, water circulation and ecosystems. As creatures evolve in worlds governed by these rules their plan-making mechanisms come to reflect them. If a particular world should on the other hand be random-based, then a creature's planning ability might produce more diversity, but there would be no logic as to when or where the plans are applied. This makes it important to measure the complexity of the world in addition to the diversity of the creatures. I propose to measure the structural complexity by the use of cellular automata [1] where, through simple interacting rules, complex overall behavior can be produced (see section 5).

The fact that there are general rules in the environment, many of which are independent of (perceived) scale, indirectly lends support for the proposed hypotheses. By adapting to – and adopting – the rules inherent in the environment, creatures might manipulate and subsequently apply rules internally (imagination) to different situations or problems. An externalization of such an internal representation would be regarded as creative behavior if a new causal relationship is discovered and applied (e.g. when a creature discovers new means to fulfill it's goal). It is safe to presume that animals do evolve consistently with these rules since failure of their adoptation would tend to prove fatal to the creatures and thereby result in the destruction of their genes.

Creativity's roots, according to this line of thought, derive from the interaction of the environmental structures and their components: If the interactions of these structures are not obvious (to the organism), perceivable and very repetitive, evolution must provide creatures with a cognitive system capable of producing diverse behavior in response to the diversity of the environment. As the unpredictability of the environment increases, or in other words as the interactions become so complex that causation is perceptually invisible, the creatures must evolve an ability to mentally represent and link unobvious causations. Another reason might be that visual cues appear once it is too late for the creatures to react, this would also support the evolution of a mechanism capable of visualizing possible scenarios before they actually occur. The mechanisms of creativity must therefore be adept at building mental imagery – and in parallel keeping the imagery logical and relative to the environment.

This co-evolution of creativity and external environmental logic could result in internal visualizations arranging in components and structures corresponding to the structural rules of the environment.

In the present framework the creatures are still at the evolutionary stage of relative reactiveness – they depend solely on visual cues in the environment to initiate a plan. However, the plans themselves directly serve as a prediction of the environment and are sufficient to produce results which support or refute the main hypotheses. [2]

It could be argued that this definition of creativity is flawed since the concept of creativity is commonly used as a measurement of artistry: Artists are creative, everyone else is not. This, however, is not a scientific definition (and can arguably be proven wrong). In the case of creating plans, a creature's body simply replaces the paintbrush; whether it is to express a certain feeling in the form of a painting - or a plan intended to reach food, every externalization of thoughts has its roots in a cognitive system – the ability to vary and adapt these expressions in response to a particular situation is my definition of creativity. Therefore I believe that the common definition of creativity, such as making music, is the same ability as evident in diverse behavior of "inferior" animals.[3]

Following the terminology introduced by Holland [6], covering natural and artificial adaptive systems, a gene is denoted A, composed of set of alleles is $A = \{a_1, a_2, \ldots a_n\}$. In the example in this paper, A denotes a unique component within the genetic structure to which genetic operators, Ω, are applied. I call these structures "control boxes". In my work the structures corresponding to Holland's alleles are connections between these control boxes. The combinations of control boxes and their connections are defined by the adaptive plan τ which uses the genetic operators. A listing of the genetic operators can be found below on the Example Genetic System (section 5.2). The criterion for comparison of τ will is denoted as μ. The measurement when comparing adaptive plans is the age a creature reaches by the use of a particular adaptive

[2] The creatures will execute a plan in order to acquire food; to serve it's goal, the plan must be in accordance to the environment.

[3] Although the purpose of this paper is not to answer the question of where or why the human superior level of creativity originated, we might postulate that the human physical embodiment (e.g. stereo vision, opposable thumb) interactively resulted in an evolutionary path emphasizing cognitive machinery capable of such creative behaviors, and that other animals such as bees or elephants were less fortunate.

plan. The set of structures attainable by the adaptive plan can represented by the following equation:

$$a = A_1 \times A_2 \times \dots A_n = \Pi^n_{i=1} A_i \quad \textbf{(1)}$$

Since the environment, E, selects over time which control boxes survive between generations, an evolving, adaptive plan is defined over time by a particular environment E.

4 The World

Cellular automata has been used frequently in the studies of A.I. under the category of ALife (Artificial life), where the cells are programmed to organize themselves to form patterns which behave like creatures [7]. This framework takes a different approach as the creatures that inhabit the world have pre-programmed cognitive systems. The worlds that the creatures inhabit consist of a 2-D grid of colored 'patches' or squares. The patches represent physical phenomena such as grass and rocks. At regular intervals a particular environment changes – grass grows and dies, etc. This is accomplished with the use of cellular automata [1]. The term refers discrete dynamical systems in which patches or cells on a grid layout are given local rules to abide. These rules provide instructions as to what state the cell should take depending on its neighboring cells. One important aspect of using cellular automata in this framework as the basis for the virtual worlds is that the patterns that the environment will be based on regularities; even though the rules of each type of patch are unknown to an observer there is a visual, underlying structure in the general behavior. Another even more important feature is that the rules are explicitly represented.

Highly complex behavior can be produced by the use of simple rules while at the same time keeping the world logical in terms of general behavior, since the state of an individual cell is dependent on its surrounding neighbors. By varying the rules and number of different kinds of patches, the diversity of the environment can easily be measured since the variation of patches and the number of their rules constitute the worlds complexity/diversity.

A number of unique worlds are proposed as testing grounds, two of which can be seen below on Fig 1. To provide a sensible measure as to how the diversity affects plan composition of the creatures, an appropriate balance between the quantity of food in each type of world must be kept. This will ensure that environmental niches [6] are kept at a minimum, giving the different worlds equality in everything besides diversity.

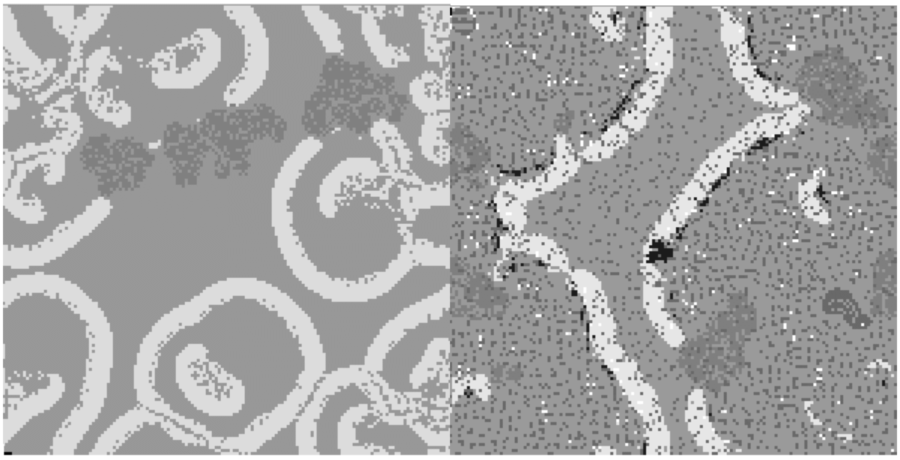

Fig. 1. Examples of environments. A Simple World (Es, left) and a Complex World (Ec, right). The difference in complexity is computed by analyzing the cellular automata rules that govern the worlds along with the number of visually distinct states.

5 Creatures

The artificial species should be designed to bear a resemblance to natural insects found on Earth, this information is based on the research of Collet & Collet [8]. The simulation of insect memory and navigational abilities has been attempted using a physical LEGO robot [9]. The authors showed a method for reducing the degree of prior knowledge required about the environment, compared to other more classical approaches. The learning method they used was comparable to the one proposed here: Insects associate visual cues with their plans to navigate to and from nest and food sources. The method for associative learning and cognitive apparatus used here is a very simplified version of the one applied in the LEGO robot.

In my framework a prototype (in development) has a body comprising a vision system, mobility, a mouth for feeding and a mental apparatus (Fig 2). A memory system provides the basis for forming repeatable plans that enable them to interact with their environment. The cognitive mechanisms, (labeled 1-7 on Fig 2) are divided into parts as described in the next six sections. The parts of the creature are:

1. Perception Cortex
2. Decision Cortex
3. Episodic Memory
4. Genetic System
5. Plan Composer
6. Motor Cortex
7. Digestive System

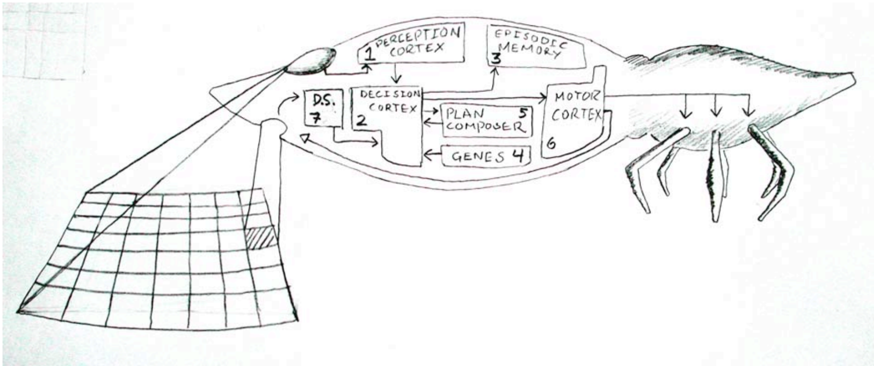

Fig. 2. The creature, showing its main functional units, along with the visual field and edible patch. (D.S. = digestive system.)

We will now look at two examples of how components in a creature could be implemented. These are not part of the platform itself but rather are described here to give an idea of the level of complexity and functionality of these key components.

5.1 Example Episodic Memory

The Episodic Memory (3, Fig 2) is the storage area for all perceptual input, the creature's actions and the consequences of those actions. In my example, memories are stored in the form of chunks which integrate these various components together into a package. The packages include a particular visual field, the plan that was associated with the perceptual input, the score - in particular the energy expenditure which was registered during the execution of the plan, along with a normalized plan score.

During each step, the Decision Cortex sends a message to the Episodic Memory to find all memories that the creature might have of similar situations. The procedure is as follows:

Find all similar memories by first rating them along three semi-independent feature dimensions:

1. **Image Matching** (dimension one). The visual field in general: Compare each patch of the current visual field to each patch of a particular memory M. Similarity scoring is incremented with each patch that corresponds precisely to a patch in the precisely same location in memory M, based on the creature's coordinates (thus, if a patch in the upper left-hand corner is green, and the upper-left hand patch in memory M is green, the memory gets a point for similarity). This score is normalized, so the increment for each patch that's identical is 1 divided by the total number of patches in the visual field. The final score is the sum of all identical patches

2. **Dominant Zone Similarity** (dimension two): Compare the current visual fields dominant color and the zone the dominant color is mostly evident in to all memories. The score is normalized, but is either on or off (i.e. 1.0 or 0.0).

3. **Patch in Front of Mouth** (dimension three): Compare the patch in front of the creature's mouth to that same patch in all existing memories. This is similar to the

Zone Similarity scoring -- if the current patch in front of the mouth corresponds to that same patch in a memory, the similarity is turned on (again, 1.0 or 0.0).

For each of these features, the memories are arranged in descending order according to each feature's similarity score. The plan associated with each memory (there is either a plan or a "primitive move" associated with each memory) is retrieved for the top three memories -- one for each dimension.

Plan scores are normalized by the following equation:

$$S_f = 1 / Sp(S_{max} - S_{min}) \quad \textbf{(2)}$$

where S_f is final score, Sp is plan score, S_{max} is the maximum score of any plan retrieved from memory, and S_{min} is the minimum score for any plan retrieved from memory.

The normalized plan scores of memories in the sorted dimensional arrays are multiplied by the similarity measurement of the memories, providing a "winner" which then determines which plan is used.

5.2 Example Genetic System

The Gene Manager (4, Fig. 2) is the part of the program that gets passed on and mutated between generations by Ω (genetic operators); it holds the creature genome which constitutes a set of control boxes and the connections between them (Fig. 3), these two components constitute the Control Network. The control boxes consist of two inputs I1 and I2, an output O and an operator which determines what to do with the inputs received. There are no initial connections between control boxes for the first generation in all worlds, and the initial outputs of Static Boxes are set to zero. After the first generation, all settings of control boxes and their connections are changed by mutation. The types of boxes in the Control Network are:

Static Boxes have a fixed output that stays constant throughout the lifetime of an individual.

NOR Boxes or “Not OR” boxes return 1 if both inputs are zero.

NAND Boxes or “Not AND” boxes always return 1 except if both inputs receive numbers.

Input Gate Boxes return I1 as output only if the value of I2 is more than zero.

5.2.1 Control Network

Each of the boxes serves as a node in a Control Network (Fig. 3). Outputs from the boxes are connected to none, some or all of the other box inputs. This is implemented by using a matrix to represent the outputs and inputs of all of the boxes (Static Box inputs excluded). The exact routing of the control box inputs and outputs is determined entirely during reproduction of the creatures. That is, the creatures’ genetic configuration is responsible for utilizing the inputs received efficiently for plan composition.

For each box, a total of 4 possible connection configurations are possible. Given an example setting of 39 boxes that each can receive two inputs (static boxes have no input) and provide one output, the number of possible connection structures attainable by the network are 4^{39}. An interesting question is the relationship between the this number and the minimum number of generations required to develop sustained existance. An even more interesting question is the relationship between the complexity of the world and the size of the Control Network.

On every step during execution the Control Network receives a request from the Decision Cortex for updating its outputs. The outputs feed into the Plan Composer and set control parameters as the creature decides how to plan its actions. There are five main ways for planning the next step, as explained in the next section.

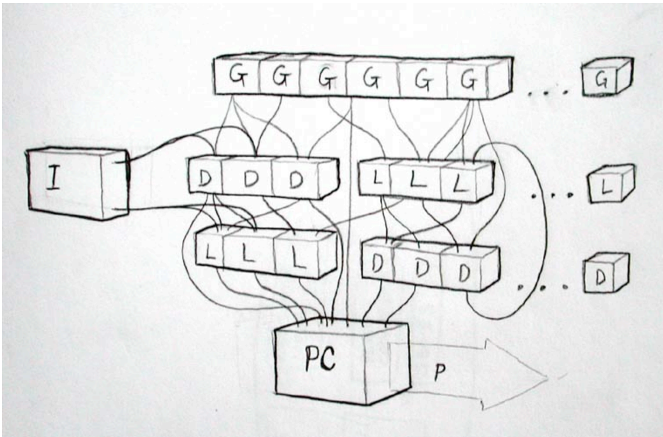

Fig. 3. Control Network. Output flows from the bottom of the boxes into none, some or all inputs of the other boxes (except for Static Boxes, see text). Boxes can connect to themselves. (G = Static Box, D = NOR Box, L = NAND and Input Gate Boxes, I = inputs from perception and internal state, PC = Plan Composer, P = plans.)

5.3 Example Plan Composer

The Plan Composer (5, Fig 3) receives instructions from the Control Network during each turn regarding what to do and specifically how to do it. The instructions are in the form of integers that are deciphered by simple logic gates applied to the Control Network inputs in the Plan Composer. The methods that the Plan Composer can use to create the next plan include:

a. Create a new plan from scratch**.**

b. Combine halves of two old plans.

c. Sequenced composition (Combine two whole plans, second plan executed after the first).

d. Mutate an old plan. (Randomly change primitive actions of an older plan)

e. Use an old plan unmodified.

When creating a new plan from scratch, the Plan Composer randomly selects the length of the plan and how many instances of each primitive action are to reside in it. With random distributions of actions, the other four methods become very important since they provide a much more controllable way of making sensible plans. Also, by using combine and mutate, the individual development will become more evident as the creature is bound to use the methods on plans that have provided good results.

6 Quantifying Creativity and Complexity

6.1 Goal-directed creativity as especially evident in humans is generally very hard to measure since it is expressed in such highly diverse manners. By giving simulated creatures the single goal of surviving, the logicality of new behavior can be directly related to that particular goal, and hence, be measurable as longevity of the creatures. The number of distinct plans in memory at the time of death, produced to fulfill that goal, can then be measured to quantify creativity.

According to the hypothesis that creativity is governed by environmental diversity, creatures in random environments (or extremely complex) should evolve to become overly creative, creating plans that have little or no persistence between generations when it comes to the number and distribution of primitive actions (because there is no logical persistent structure in the environment); whereas in dynamic logical environments the creatures should reach a sustainable balance in primitive action usage. The randomness of the environment will therefore be reflected by a disproportion of primitive actions in the creatures' plans.

6.2 In my proposed method of quantifying complexity, a rule is defined as any rule or set of rules that causes a state transition for a patch under some condition or conditions. This can be taken as an estimate of the interaction complexity of a world because each rule represents a potential for interaction. The world's complexity is quantified by counting the number of (perceptually distinct) components and the number of rules these components abide. These numbers provide a ratio that represents the relative creativity over the complexity of the world. A resulting *Complexity Quotient* is defined as:

$$Qc = Pi * C \quad (3)$$

where C represents the number of components (patches) perceivable by the creatures and Pi represents the number of rules pertaining to these patches. Rule examples can be found below in Table 1.

The ratio between environmental complexity and mental creativity is proposed as a mathematical definition of creativity. This definition assumes that the creatures' perceptual abilities are held constant. A more complex measure could be proposed (see Discussion).

Table 1. Rules of example worlds on Figure 1, Simple (Es), and Complex (Ec)

Simple World (Es; Qc = 18)	**Complex World** (Ec; Qc = 91)
-If eaten: turn brown -If green and there are more than 20 green patches around and lifetime exceeds 30: turn brown -If green and there are less than 12 green patches around and lifetime exceeds 20:	- If green and lifetime exceeds 53: turn dark brown - If green and the number of green patches around you are more than 23 and lifetimes exceeds 30: turn magenta. -if green, the number of green patches

turn brown -If green and number of green patches around equal 25, turn brown. -If green and lifetime under any circumstance exceeds 60: turn brown - If brown and there are more than 8green patches around, their lifetime combined exceeds 80 and there are more than 10 brown patches around: turn green.	around you is less than 12 and lifetime exceeds 20: turn brown. - if green and there are more than 6 red patches around: turn red. - If magenta and lifetime exceeds 30: turn yellow. - If magenta, lifetime is less than 30 and the number of brown patches around you are more than 19: turn green. - If brown, the number of green patches are more than 6, number of brown patches are more than 10 and lifetime exceeds 25: turn green. - If dark brown and the number of magenta patches are more than 4: turn yellow. - if dark brown, the number of green patches are more than 6 and the number of dark brown patches are more than 4: turn red. - If red and lifetime exceeds 15: turn brown. - If yellow and counter of life exceeds 30: turn magenta. - If yellow, lifetime is less than 50 and the number of magenta patches is more than 8: turn brown. - If yellow, lifetime exceeds 70 and the number of brown patches is more than 19: turn green.

7 Discussion & Future Work

The potential of an autonomous system to act intelligently in the world is largely due to its ability to compose new ideas and propose solutions. Essential to this type of behavior are the creative mechanisms, the ability to discover new concepts or relationships in the environment and apply them in a goal-directed manner. This paper describes ongoing work that looks at the creative capabilities of an evolving cognitive system and describes the foundation for further analysis. The platform I propose can be used to simulate the evolution of cognitive systems capable of expressing their creativity in the form of plans. By building a simulation, psychological and environmental variables can be quantified and are under direct control.

The formula for measuring the complexity of the environment, Qc, shows promise for quantifying this feature and its simplicity is an obvious benefit. However, it might turn out to be insufficient since the expression of rules is time-dependent. I intend to improve the formula in the future by taking into consideration the amount of time a particular world exists. Normalization of the scale is also a future task. Another question regards whether Qc should incorporate the creatures' standpoint by calculating the perception space of the creatures – in other words how much of the environment it is possible for the creatures to see considering her initial energy and/or perceptual abilities. This will be explored in the future.

Follow-up research might involve varying the expressional abilities of the creatures: giving them new means of manipulating the environment (other than just eating it). Another possibility is to enhance their perceptual abilities – e.g. giving the creatures a sense of smell or hearing.

The framework should produce interesting empirical results in the near future. Software will be made available for anyone who is interested in using it for their own work.

Acknowledgements

I would like to thank the following individuals and companies for their support and encouragement: My brother Dr. Kristinn R. Thórisson for continued support, advice and technical assistance. My fiancée for her patience and understanding. My father, mother, other brother, sister and friends for moral support. Ólafur H. Sigurjónsson – vice principle of the Junior College in Ármúli, Stefanía Kristinsdóttir, the University of Iceland, Apple IMC Iceland, Mayor Jónmundur Guðmarsson and the town-council of Seltjarnarnes, Þórólfur Árnason the mayor of Reykjavík, and the anonymous reviewers for their comments.

References

1. Wolfram, S.: A New Kind of Science. Champain, IL: Wolfram Media. (2002)
2. Carruthers, Peter.: Human creativity: its evolution, its cognitive basis, and its connections with childhood pretence. British Journal for the Philosophy of Science 53. (2002) Retrieved May 19, 2004, http://www.philosophy.umd.edu/people/faculty/pcarruthers/Creative-thinking.htm
3. Nils A. Baas and Claus Emmeche.: "On Emergence and Explanation", Intellectica, no.25, (1997) p.67-83. Retrieved June 28, 2004, http://www.utc.fr/arco/publications/ intellectica/n25/25_04_Baas.pdf
4. Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., & Yuste, R.: Synfire chains and cortical songs: Temporal modules of cortical activity. Science, 304, (2004). 559-564.
5. Thaler, S. L.: The machine that invents, Post-Dispatch, 25 Jan. (2004) Retrieved June 22, 2004, http://www-acad.sheridanc.on.ca/staff/eagen/Notes/Creativity Machine.pdf
6. Holland, J.: Adaptation in Natural and Artificial Systems. Cambridge, MA: MIT Press. (1998).
7. Antony, A., C. Salzberg, H. Sayama,: Evolutionary dynamics of cellular automata-based self-replicators in hostile environments, BioSystems, preprint (accepted pending revisions).
8. Collet, T. S. & M. Collet.: Memory Use in Insect Visual Navigation. Nature Review, July, Volume 3. (2002)
9. Chan, P. and G. Wyeth.: Self-learning visual path recognition. In: Proceedings of the Australian Conference on Robotics and Automation, (1999) pp. 44-49, Brisbane

Automated Control of Interactions in Virtual Spaces: a Useful Task for Exploratory Creativity

Federico Peinado, Miguel Ancochea and Pablo Gervás

Dep. Sistemas Informáticos y Programación
Universidad Complutense de Madrid, Spain
fpeinado@fdi.ucm.es mancochea@indra.es pgervas@sip.ucm.es
http:\\gaia.sip.ucm.es

Abstract. The main problem for Interactive Digital Storytelling is the inevitable conflict between author's determinism and interactor's freedom. This paper presents a computational solution for addressing this problem, and discusses how exploratory creativity (more than transformational creativity) can play a role in solving the dilemma. Our proposal is based on the theoretical study of tabletop Role-Playing Games, Virtual Environments applications and structural analysis of narrative texts. It involves the implementation of automatic directors for the *story* and *discourse* levels useful for many story-based applications.

1 Introduction

The use of physical metaphors to help with the organization of information presents important advantages from the point of view of navigability, ease of access to particular items of information, and orientation of the user while exploring a particular set of data. As more and more institutions become virtual (retail outlets for products such as books or computers, museums, universities, banks, libraries, newspapers...) this type of solution for presenting information is becoming increasingly popular. Even if it is not made explicit, a hypertext web site can be understood as a virtual space if particular documents are interpreted as rooms and hyperlinks as connections between them.

Designing a virtual space involves finding a trade-off between two possibly divergent aims: to include enough similarity to an existing concept to act as guidance for the user, and to exploit the advantages of the virtual space over the concepts which it mirrors (every available object can be as close as the next room, experiences can be explicitly tailored to each individual, displays can be dynamically rearranged based on user feedback...). In many cases, designers of web sites tend to focus on exploiting the advantages of the virtual space. This leads to web sites where the user can easily become lost, overwhelmed by the amount of information, or unable to find something that he knows must be somewhere close. As a result, several research initiatives have addressed the problems of guiding the user around this type of virtual space, particularly centered around electronic catalogues and virtual museums [7,15,18,21].

An extreme case of this type of situation is provided by the emerging field of Interactive Digital Storytelling, in which the concept of virtual space is blended with literary fiction. The main problem of Interactive Digital Storytelling (IDS) design is structuring an intensive and meaningful interactive experience at the same time as enabling the development of a good pre-authored plot. This is a real dilemma, because while the interactor of some story-based application is taking a lot of decisions about the performance of his character that change the plot development, the author's plan may have been developed according to a different storyline.

This paper presents a computational solution for addressing this problem, and discusses how exploratory creativity (more than transformational creativity) can play a role in solving the dilemma.

2 Relevant Techniques and Useful Sources

Many approaches are found in the literature that try to solve this conflict in an automated or semi-automated way. Basically, they make interactive storylines by adapting author's plot units or other pre-established resources to the interactor behaviour at run-time. This task requires a computational solution that can react appropriately to "unexpected" user decisions.

In order to do this, several aspects have to be taken into account: some guidance as to how the control of the interaction may be bettered achieved, a valid heuristic for finding new solutions based on existing ones, appropriate technology for rendering the messages to the user as natural language text, and an adequate representation of the needs or preferences of particular users. This section covers the technologies and sources used in this work to cover those aspects, as well as presenting an existing Interactive Fiction engine that is used as case study platform to test the solutions presented here.

2.1 Basic Source for Control Rules

Because Interactive Narrative is a relatively new field and it is difficult to find formal studies about interactive plot development, we have done a review over the modern methodology of Role-Playing Games (RPGs), the interaction in Virtual Environments (VEs) and the classic studies about narrative structuralism like *Story and Discourse* by Seymour Chatman [5].

Tabletop RPGs are exercises in intellect and imagination: a group of players sitting around a table, rolling dices and playing out an imaginary role in a complex shared fantasy, true collaborative narrative.

The Game Master (GM) is a special kind of player, he is the "interactive storyteller". He designs all the elements of the story and he manage all the possible events that can occur in its development, improvising the dialogue contributions of non-player characters, resolving players actions, etc.

The degree of interactivity in RPG can be enormous, limited only by the players imagination. This implies that no GM, however experienced, can have a deep enough plan to react appropriately to all the possible actions that players can come up with in the world of fiction. To operate successfully without such a plan, GMs must use their imagination, improvise adequate solutions, and continuously rewrite their plots on the fly.

The figure of GM is the best model we have found in real life for designing and directing interactive stories. For the development of the work presented here we have used the description of the relevant heuristics given by Robin Laws [13].

2.2 Knowledge-Intensive Case-Based Reasoning

Knowledge Intensive Case-based reasoning relies on applying additional explicit knowledge to improve the performance of case-based reasonings approaches that rely mostly in reusing existing solutions to adapt to new problems. The COLIBRI (*Cases and Ontology Libraries Integration for Building Reasoning Infrastructures*) system assists during the design of KI-CBR systems that combine cases with various knowledge types and reasoning methods. It is based on CBROnto [8,9,10], an ontology that incorporates reusable CBR knowledge and serves as a domain-independent framework to develop CBR systems based on generic components like domain ontologies and Problem Solving Methods (PSMs).

2.3 Natural Language Generation Architecture

FROGS is a flexible object-oriented Java-based framework to build Natural Language Generation (NLG) applications taking RAGS [3] as a reference and thus implementing its main standard definitions for the abstract data model and using XML for the real source data.

This tool, which also provides a sample default implementation called *jGolen*, supports a wide selection of the most common state-of-the-art generation architectures, ranging from a simple monolithic implementation to a revision-based architecture or a blackboard and including the frequently used Reiter's pipeline [23] or the interactive feedback-based architecture.

These architectures usually carry out a series of common generation tasks also provided by the framework, including content determination, discourse planning, sentence aggregation, lexicalization or linguistic realization.

2.4 User Modeling

In order to apply the personalization described above, it is crucial to identify correctly the character profile of the interactor.To simulate this dynamically in an interactive system is the next step of this study.

Interactor Models for Story-Based Games In a RPG, identifying the character profile of the interactor must be carried out based on limited information like the description of the character explicitly required by the rules, observation of the player's reactions, and possibly the player's record in previous games. For our current purposes, it is enough to provide interactors with a set of predefined characters, such that each one of them is related with a specific interactor model. It is hoped that interactors of a specific model will under such conditions choose the type of character most related to their preferences in acting. The initial adscription of interactor model to character type will be used by the system to assign interactor models to the interactors.

Laws identifies seven basic types of role-players according to their motivation and the sort of characteristics that they expect of a game in order to consider satisfactory. These motivation characteristics are referred to as *plot hooks*.

Power Gamer searches for new abilities and special equipment.
Butt-Kicker always waits for a chance to fight with someone else.
Tactician feels happy in the anticlimax scenes, thinking about logical obstacles to overcome.
Specialist just wants to do the things that his favorite character do.
Method Actor needs situations that test his personality traits.
Storyteller looks for complex plot threads and action movement.
Casual Gamer remains in the background (and has a very low degree of participation in general).

Interactor Models for Virtual Museums It is widely accepted and many authors agree [21,18,15,17] that moving on a virtual space or environment as a virtual multimedia-based museum needs user modeling in order to provide the most convenient adaptation mechanisms to make every visitor enjoy a so-called "museum experience".

Keeping the visitor interested during all the route is essential, as it is also giving him/her the most suitable information, so that it is not repeated nor more than needed and coherent with the previously given information.

Aiming to model the museum visitor as closely to the real thing as possible there have been a series of initiatives in many fields to determine what are the main behavioural features to be considered of a person in that circumstances, just as it has been modeled on Role-Playing Games. In the HIPS approach [21], the authors take into a count a very useful classification based on previous psycho-social studies.

Ant visitor tends to follow the path proposed by the curator.
Fish visitor prefers to move in the centre of each room, without looking at details of the art, but cruising all the exhibition.
Butterfly visitor frequently changes direction, and don't follow the purposed path, although they manage to see almost all the artworks.
Grasshopper visitor sees only the pieces of artwork they are interested in, leaded by their personal interests and knowledge.

2.5 Internet Adventure Game Engine: A Case Study

Text Adventure Games, broadly known as the Interactive Fiction genre, appeared as the first narrative games at the end of 70's. Originally, interactive fictions are like interactive books, only made of text chunks. They have complex plots and offers a narrative presentation to the player. In this kind of applications, story and world simulation are tightly coupled.

Internet Adventure Game Engine (IAGE [22]) is a Java Open Source project for the development of a multiplayer interactive fiction system. In contrast to Massive Multiplayer Online Role-Playing Games (MMORPGs), which maintain a lot of players playing at the same time, with as many ongoing stories as users connected to the server, IAGE allows one pre-authored storyline with the added possibility of having more than one player immersed in the same story. IAGE can be also used to create single player applications like traditional systems as Inform [19].

3 Story and Discourse Architecture

Automated control of interactive narrative requires a system which is able to improvise acceptable and engaging solutions to conflicts arising at two different levels. One is the creation of characters, places or situations that will appear in the *story* and the other is related to the form of the messages that the interactor will send and receive during the *discourse* generation in real time (specially in case of natural language interfaces). At each level, the definition of what is considered acceptable and engaging will vary.

Our approach is based on developing a Case-Based Reasoning (CBR) model of a particular set of algorithms and heuristics for RPG mastering and VE interaction, and applying this model to narrate the experience of each interactor in a multiuser & directed i-fiction engine.

We propose a global architecture for a system which employs both CBR-driven story management, user modelling and adaptative context-sensitive discourse generation.

Firstly, there would be a *story director* as a standalone component. Like an agent *agora*, this module is intended to provide adaptative management of the VE and interactions for every "agent" on the system, including the characters and the *discourse director* components.

This last component is created once for every interactor and attached to his client application by the story director. Its aim is not only to act as an NLG module, but also to adapt generation to the interactor model and to the current state of the VE just as the interactor perceives it.

The story director performs a previous selection of the environment currently available for each interactor, as a graph of facts and entities, and gives it to discourse directors, enriched with the history of previous interactions. Discourse directors should then generate texts which try to attract and interest the interactor, at least as the system can see him.

There would also be modules in charge of user input processing or interpreters, with raw natural language processing capabilities.

As a technological standard of reference to build a system based on this architecture, we propose that of a J2EE multitier application. This way, there

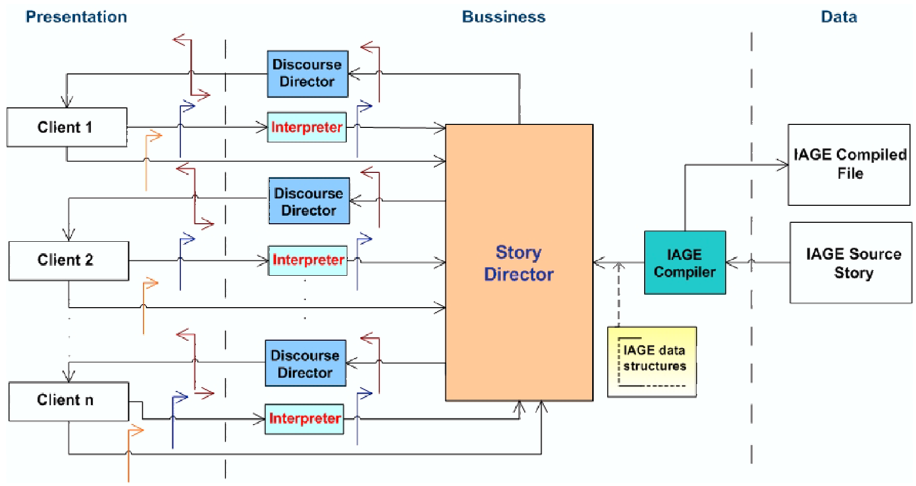

Fig. 1. Overview of the multitier client/server IDS system architecture

would be a lightweight-client presentation layer, a bussiness server layer involving all story an discourse management processes and a back-end server data layer, as shown on Figure 1.

Interfaces are shown as arrows pointing in the direction of the necessary implementor. Upon execution of a lightweight client application, it registers itself on the story director, which in turn instantiates a discourse director and an interpreter and attaches them to the client application, so that it only communicates with them from then on. Each interpreter processes input, passes it to the story director and it decides whether to notify to the corresponding discourse director any relevant change on the VE or interactor, thus implementing an architectural Observer pattern, on which the story director is the subject to be observed. In order to decouple the parts of the system involved in this communication, an additional Factory pattern should be employed to instantiate discourse directors for the story director.

3.1 Case-Based Reasoning Story Director

In searching for the right computational solution, widely different approaches have been proposed by researchers in the area. For example, there are proposals based on dynamic behavior of autonomous characters that achieve dramatic goals [4,25]. Other approaches give more responsibility to a central dramatic planning algorithm, using directable characters [14,11] or just a stand-alone dramatic planner that controls the most important narrative elements, like characters [16] or the whole fictional world [12]. In the work of [11] the CBR full life cycle -retrieval, adaptation, reuse and repair of previous solutions for new problems- is used for storyline representation and a strategy formalization that allows for storyline adaptation.

We propose a Knowledge Intensive CBR (KI-CBR) approach to the problem of generating interactive stories from the game state and a set of cases specifying how the next scene will be performed. In our model of Interactive Storytelling, adapted from the original RPG conventions, we separate the world simulation from the story control. The IDS system that we are considering has a narrative environment with a high level of interactivity (via textual commands) and uses IAGE as a low level world simulator. Over that, a CBR system guides the development of plot, building creative new situations on the fly from the case base.

Knowlege Representation In order to be able to use the Chatman's concepts and the user models computationally, they have been translated into an ontology that gives semantic coherence and structure to our case base.

An additional ontology provides the background knowledge required by the system, as well as the respective information about characters, places and objects of our world. This is used to measure the semantical distance between similar cases or situations, and maintaining a independent story structure from the simulated world.

The author fills the case base with scenes that contain preconditions and postconditions. Because the case base is made using cases proposed by the author, we know that the system makes variants of many elements that can be combined in a lot of different ways.

Plot Driving Algorithm Our plot driving algorithm is based on a CBR interpretation of the Law's improvising method [13]. This method is based on making transitions to the most *interesting* story scenes whenever it is possible.

The CBR system uses two similarity functions. The first one is used to recover the scene that leads to the most obvious/surprising transition from the current scene. Every case has a obvious/surprising property and the author assigns negative values to obvious scenes and positive values to surprising ones.

The second similarity function is used to retrieve the scene that involves the most pleasing/challenging transition from the current scene. The definition of *how pleasing/challenging a scene is* is given by the number of easy/difficult tasks matching the interactor plot hooks that appears in the scene. The author selects one kind of interactor as a reference and assigns negative values to pleasing scenes and positive values to challenging ones, based on those things that the interactor prefers.

For example, an ambush during a walk in the Palace's garden is a surprising scene, a big challenge for a Butt-Kicker playing an adventure game. On the other hand, free exploration of the museum shop at the end of the visit is an obvious but pleasing transition for a Butterfly visitor.

In this way the algorithm includes a number of obvious paths and other paths that may progressively get more positive for the interactor interests. Additionally, it may include surprising or negative scenes.

3.2 Natural Language Discourse Director

On a general, architectural basis, it is a relatively complex module of adaptative generation. Its core is intended to be standalone and not bound to a domain or the global system on which it is used, making therefore necessary a domain adapter or wrapper. This adapter processes its domain-dependent input and builds a series of generation and user model hints to be passed to the core component.

The core itself consists of a component of raw language generation and a module to adapt the previous hints to the current user an environment context, a contextual adapter. Eventually, this new adapter will start generation with a series of hints and provide a textual planning algorithm by building a text-potential graph based on entities, facts and relations, similar to that found on ILEX [17], so that the core NLG plans easily by implementing a traverse algorithm for the graph.

Finally, the NLG core is intended to be an implementation of the FROGS framework, thus implementing also the standards proposed in RAGS as a reference architecture [3,2].

4 Discussion

In the present context, when we are considering the automation of the task of dynamically controlling the flow of interactive narrative, it becomes imperative to discuss at least two basic issues:

- whether this task (or parts of it) indeed involves creativity at some level
- how this particular type of creativity can be classified in terms of Boden's analysis [1]
- whether this kind of creativity can be automated in any way in a system of the kind described

It is clear that controlling interactive narrative, such as acting as Game Master in a Role Playing Game, is generally perceived as a creative task. Human candidates that carry out the task have to fulfil certain expectations on the part of the players, and knowledgeable players are quick to judge whether a game master is good or not. However, there is no easy way of extrapolating explicit rules as to how this evaluation is carried out.

On the other hand, the generation of literary fiction, even of the non-interactive variety, is considered creative. Art and music require the generation of artefacts that are radically creative, in the sense that they cannot be classified under existing genres. In these fields, an artefact is deemed creative only if it defines a style of its own. In the field of literature, the situation is more complex. The definition of a new genre is not an immediate aim of the community of creators. Rather, they tend to focus on being creative within given genres.

One possible explanation of this difference is related to the role of meaning in each of these fields. Music, and non figurative art, produce artefacts that have no intrinsic meaning. In literature, on the other hand, meaning is crucial, and it probably gives rise to a very high percentage of the value of the artefact. This is not the case in certain types of free style poetry.

Another possible factor to take into account is utility. Wherever artefacts are expected to be useful, or to fulfil a specific set of needs, the kind (or the degree) of creativity expected is reduced. In a way, a particular set of needs to be fulfilled constitutes a skeleton definition of a genre.

This has become particularly apparent with the mass production and dissemination of musical and literary material. People have come to expect certain artefacts to fill certain requirements. For instance, airport novels must grip the reader and keep him entertained over a reasonably long period while requiring little effort. Pop songs intended to be played in radio stations are expected to have a certain length, to be catchy, and to have refrains that get repeated a few times. In either case, there is an industry that sets the requirements and drives the type of material that is produced. Should this be taken to imply that no creativity is involved? This is not true. In fact, within each field, success depends very much in being creative within the given requirements, as this becomes the distinguishing feature of a particular product with respect to its competitors.

However, it could be interpreted to mean that the kind of creativity required in these cases is not transformational creativity, but rather exploratory creativity. In general, it is easy to assume that the amount of creativity required to produce valid artefacts in these fields is very low. We have all heard someone complain about the injustice of bestselling authors who write books by the rules and make millions out of them, having exercised very little "creativity". If this were true, it ought to be easy to program computers to carry out this kind of tasks. We have yet to hear about major breakthroughs of computer authored artefacts in any of these industries.

From the point of view of the feasibility of automating in any way the creativity involved in these tasks, it is important to take into account that the kind of creativity that is being sought is not transformational creativity but exploratory creativity. In the particular domain of interactive fiction, and narrowing down to the proposed case study, the amount of creativity required would be restricted to finding adequate combinations of whatever resources (locations, characters, objects, events, character functions...) are explicitly represented in the available ontology.

The results obtained by the system when carrying out the task of controlling real interactions can be evaluated in two ways. On one hand, they can be compared with the results of equivalent tasks carried out by human operators. To this end it is important to provide the system with adequate facilities for keeping logs of particular games. This evaluation needs the intervention of human users. On the other hand, attempts can be made to evaluate from a more formal perspective the level of creativity actually displayed by the system. This evaluation can be oriented towards locating any indications of creativity introduced by the process in comparison with the original samples which make up the case base, possibly by applying metrics and analyses of creative activities that have progressively emerged over the recent years [24,6,20].

5 Conclusions

The task described in this paper constitutes a good example of a circumstance where exploratory creativity appears not as a “lesser sister” of transformational creativity but as a more adequate alternative. The actual requirements imposed by the task suggest that a solution based on exploratory creativity rather than transformational creativity would be more adequate, and that the level of creativity required may be susceptible of automation to a certain degree. As in many cases where creativity is required, some degree of aesthetical value is needed for the results obtained to be considered acceptable. However, in the context of interactive narrative, the level of aesthetical quality from a literary point of view is quite low when compared to narrative in general fiction. Consumers of this type of narrative tend to focus more on issues like the sequence of events or the pace of the interaction than the aesthetic quality of system messages from a literary point of view. This presents many advantages since even results of low aestehtic quality (as may be obtained in the early stages of the development of the system) may still be valuable from the point of view of interactive narrative.

Although the system is not fully-implemented yet, the progress so far points to a reasonable solution for interactive narrative generation. As outlined in the introduction, once the system is fully developed, the approach presented here for interactive narrative may be extended to other situations where automated control of interactions in a virtual space is required. The design of the system has taken this possibility into account, by ensuring that adequate modularity is included in the design. In this way, domain specific resources such as user models are isolated in particular points of the code to allow for easy interchangeability.

As possible future applications of the resulting system techniques, we are considering two areas that involve interaction in virtual spaces for which specific developments are currently under way in our department. One possibility is a Virtual Museum of Computer Science that is under construction in our university. This project involves the creation of a large virtual space in which various materials concerning computer science will be displayed. Our research group is keeping in contact with the museum’s development team with a view towards ensuring compatibility in data formats. The aim is to develop an automatic museum guide that would monitor the user’s navigation through the museum, providing customized generated comments and dynamically redesigning the floor plan to suit user needs. The other possibility concerns JV2M, a knowledge-based learning environments environment where students can learn the Java Virtual Machine (JVM) structure and Java language compilation. The system presents a metaphorical 3D virtual environment which simulates the JVM and the user is symbolized as an avatar which is used to interact with the virtual objects. An animated pedagogical agent called Javy (JavA taught VirtuallY) monitors the student whilst she is solving a problem with the purpose of detecting the errors she makes in order to give her advice or guidance. The current version of the JV2M system would clearly be enhanced by the application of the techniques considered here.

6 Acknowledgements

The first author is supported by a FPI Predoctoral Grant from Complutense University of Madrid. The work was partially funded by the Spanish Committee of Science & Technology (TIC2002-01961).

References

1. M. A. Boden. *The Creative Mind: Myths and Mechanisms*. Wedenfield and Nicholson Ltd, London, 1990.
2. L. Cahill, J. Carroll, R. Evans, D. Paiva, R. Power, D. Scott, and K. Van Deemter. From rags to riches: Exploiting the potential of a flexible generation architecture. In *Association for Computational Linguistics*, Toulouse, France, 2001.
3. L. Cahill, C. Doran, R. Evans, C. Mellish, D. Paiva, M. Reape, D. Scott, and T. N. In search of a reference architecture for nlg systems. In *7th European Workshop on Natural Language Generation*, pages 77–85, 1999.
4. M. Cavazza, F. Charles, and S. J. Mead. Character based interactive storytelling. *IEEE Intelligent Systems*, 17:17–22, 2002.
5. S. Chatman. *Story and Discourse : Narrative Structure in Fiction and Film*. Cornell University Press, second edition, 1986.
6. S. Colton, A. Pease, and G. Ritchie. The effect of input knowledge on creativity. In A. Cardoso, C. Bento, and G. Wiggins, editors, *First Workshop on Creative Systems, International Conference of Case-Based Reasoning*, 2001.
7. R. Dale, J. Oberlander, M. Milosavljevic, and A. Knott. Integrating natural language generation and hypertext to produce dynamic documents. *Interacting with Computers*, 11(2):109–135, 1998.
8. B. Díaz-Agudo and P. A. González-Calero. An architecture for knowledge intensive CBR systems. In E. Blanzieri and L. Portinale, editors, *Advances in Case-Based Reasoning*, New York, 2000. Springer-Verlag Berlin Heidelberg.
9. B. Díaz-Agudo and P. A. González Calero. A declarative similarity framework for knowledge intensive CBR. In *International Conference on Case-Based Reasoning*. Springer-Verlag, 2001.
10. B. Díaz-Agudo and P. A. González Calero. Knowledge intensive CBR through ontologies. *Expert Update*, 6(1), 2003.
11. C. Fairclough and P. Cunningham. A multiplayer case based story engine. In *4th International Conference on Intelligent Games and Simulation*, pages 41–46. EUROSIS, 2003.
12. A. S. Gordon and N. V. Iuppa. Experience management using storyline adaptation strategies. In *First International Conference on Technologies for Digital Storytelling and Entertainment*, Darmstadt, Germany, 2003.
13. R. D. Laws. *Robin's Laws of Good Game Mastering*. Steve Jackson Games, first edition, 2002.
14. S. C. Marsella, W. L. Johnson, and L. Catherine. Interactive pedagogical drama. In C. Sierra, M. Gini, and J. S. Rosenschein, editors, *Fourth International Conference on Autonomous Agents*, pages 301–308, Barcelona, Spain, 2000. ACM Press.

15. P. Marti, A. Rizzo, L. Petroni, G. Tozzi, and M. DiligentiI. Adapting the museum: A non-intrusive user modeling approach. In *User Modelling*, 1999.
16. M. Mateas and A. Stern. Façade: An experiment in building a fully-realized interactive drama. In *Game Developers Conference, Game Design track*, 2003.
17. C. Mellish, M. ODonnell, J. Oberlander, and A. Knott. An architecture for opportunistic text generation. In *International Workshop on Natural Language Generation*, Niagra, 1998.
18. M. Milosavljevic, R. Dale, S. J. Green, C. Paris, and W. S. Virtual museums on the information superhighway: Prospects and potholes. In *Annual Conference of the International Committee for Documentation of the International Council of Museums*, 1998.
19. G. Nelson. Inform manual, 2001. http://www.inform-fiction.org/.
20. A. Pease, D. Winterstein, and S. Colton. Evaluating machine creativity. In A. Cardoso, C. Bento, and G. Wiggins, editors, *First Workshop on Creative Systems, International Conference of Case-Based Reasoning*, 2001.
21. D. Petrelli, E. Not, and M. Zancanaro. Getting engaged and getting tired: What is in a museum experience. In *User Modelling: Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*, Banff, 1999.
22. R. Rawson-Tetley. Internet adventure game engine (IAGE), 2002. http://www.ifarchive.org/if-archive/programming/iage/.
23. E. Reiter. Has a consensus nl generation architecture appeared, and is it psycholinguistically plausible? In *Seventh International Workshop on Natural Language Generation*, pages 163–170, Kennebunkport, Maine, USA, 1994.
24. G. D. Ritchie. Assessing creativity. In G. Wiggins, editor, *Symposium on Artificial Intelligence and Creativity in Arts and Sciences*, 2001.
25. R. M. Young. Notes on the use of plan structure in the creation of interactive plot. Technical Report FS-99-01, AAAI Fall Symposium on Narrative Intelligence, AAAI Press, Menlo Park, 1999.

Creative Discovery in the Lexical "Validation Gap"

Jer Hayes and Nuno Seco and Tony Veale

Dept. of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland

Abstract. Compound terms play a surprisingly key role in the organization of lexical ontologies. However, their inclusion forces one to address the issues of completeness and consistency that naturally arise from this organizational role. In this paper we show how creative exploration in the space of literal compounds can reveal not only additional compound terms to systematically balance an ontology, but can also discover new and potentially innovative concepts in their own right.

1 Introduction

Broad-coverage lexical knowledge-bases like WordNet [8] generally contain a large number of compound terms, many of which are literal in composition. These compounds are undoubtedly included for a reason, yet the idea that literal compounds might actually be essential to WordNet's usefulness may strike some as heretical on at least two fronts: first, the lexicon is a finite resource, while the space of compounds is potentially infinite; and at any rate, literal compounds can be created as needed from purely compositional principles [2]. However, these retorts are valid only if we view WordNet as a dictionary, but of course it is much more than this. WordNet is a lexical ontology, and ultimately, ontologies derive a large part of their functionality from their structure.

So, while the meaning of literal compounds like *Greek-deity* and *animal-product* may well be predictable from compositional principles alone, such concepts still serve an important organizational role in WordNet by adding much needed structure to the middle ontology. Having conceded the importance of such compounds, one is forced to address the issues of completeness and consistency that then arise from their inclusion. Completeness suggests that we strive to include as many literal compounds as are sensible, if they enhance the organization of the ontology or if there is evidence that they are in common usage in the language. Systematicity is a related issue that arises when a group of existing compounds suggests that another should also exist for the ontology to be consistent. For instance, the existence of *Greek-deity*, *Greek-alphabet* and *Hebrew-alphabet* leads to the hypothesis that *Hebrew-deity* should also exist if WordNet is to be both consistent and symmetric in its treatment of different cultural groupings.

Indeed, because literal compounds like these arise from the yoking together of two different ontological branches into one, compounding represents an important contextualization device in the design of ontologies, allowing lexical elements to be logically grouped into clusters or families that share important dimensions of meaning. This clustering facilitates both automated reasoning by machines (such as the determination of semantic similarity based on taxonomic distance) and effective browsing by humans. Sometimes this yoking results in a compound that, following Boden [1] and Wiggins [14], deserves to be called "creative", because it exhibits both novelty and value. Novelty can be measured along either a psychological or a historical dimension, while utility is a reflection of the uses to which a compound can be put. For instance, a new compound may have utility as a clustering node when added to the middle ontology if its appropriate hyponyms can be identified. Alternately, a new compound may represent an alternate nominalization of an existing concept (e.g., see Vendler's [13] insights about nominalization, and Lynott and Keane's [6] application of these insights to compound generation).

In this paper we present a process of ontological exploration to identify those areas of the lexicon that can contribute to, and may in turn benefit from, the invention of new compound terms. Since the discovery of new compound terms is essentially a process of creative exploration, we frame our discussion within the theoretical framework of creative computation. Within this framework two approaches to validating new compounds are presented: internal validation determines whether the ontology itself provides evidence for the sensibility of a new compound while external validation uses web-search to find evidence that the compound already exists outside the ontology. We then go on to show how these different strategies create a validation gap that can be exploited to identify the small number of truly creative compounds that arise.

2 Exploring the Space of LMH Concepts

Creative discovery requires that we give structure to the space of possible concepts that we plan to explore. This is made somewhat easier if we consider the meaning of conceptual structures to be grounded in a semiotic system of meaning-creating oppositions. Given a starting structure, knowledge of allowable oppositions can then be used to transform this starting point into a variety of different conceivable structures, some of which may be novel and possess value on a particular utility scale.

The notion of opposition employed here is much broader than that of antonymy. For our purposes, contextual oppositions exist between terms that compete to fill a given dimension of the same concept. For instance, *Greek*[1] and *Hindu* can each be used to differentiate the concept *deity* along a *culture* dimension, and so, in the context of *deity*, both are opposed. However, this is a contextual opposition that, unlike the role of antonymy, does not constitute part of the meaning of either concept. WordNet is a rich source of explicit antonymous oppositions,

[1] To avoid later confusion with set notion, we denote WordNet senses not as synsets but as italicized terms

but contextual oppositions must be inferred from the structure of the ontology itself and from existing compounds.

Fortunately, WordNet contains many instances of literal modifier-head terms, such as "pastry crust" and "Greek alphabet". The concepts denoted by these compound terms, or LMH concepts for short, have the lexical form M-H (such as *pizza-pie* or *prairie-dog*) and express their literality in two ways. First, they must be stored in the WordNet ontology under an existing sense of the lexeme H; for instance, *pizza-pie* is actually stored under the hypernym pie. Secondly, the gloss for the concept M-H should actually contain the lexeme M or some synonym of it. Thus, while *Greek-alphabet* is a LMH (it literally is a kind of *alphabet*, and it is literally Greek), neither *monkey-bread* (which is not literally a kind of *bread*) nor *Dutch-courage* (which is not literally *Dutch*) is a LMH concept.

2.1 A Framework for Creativity

We use the terminology of Wiggins [14] to frame our discussion of creative exploration. Wiggins, following earlier work by Boden [1], formalizes the creative exploration process using the following abstractions:

C - the realm of concepts that is being explored

R - the set of rules for forming concepts and conversely, deconstructing existing ones

T - the traversal rules that generate new concepts via R

E - the evaluation mechanism that ascribes value or utility to these new concepts

In applying these terms to creativity in WordNet, we introduce the following refinements:

C_w *- the subset of C described explicitly in WordNet as synsets*

C - the set of LMH concepts in Cw considered as a starting point for creative exploration*

R - the subset of R needed to construct and deconstruct LMH compounds in C**

T - the subset of T needed to hypothesize new LMH concepts for R* to construct*

So for our current purposes, we define C* as the set of LMH concepts in WordNet, and R* as the compositional criteria used to identify and decompose existing LMH entries and to construct new ones by concatenating an appropriate M and H term pair. However, to define T*, we first need to consider how taxonomic differentiation is used to create LMH concepts in the first place.

3 Domain Differentiation

LMH concepts exist in WordNet to differentiate more general concepts in meaningful taxonomic ways. For instance, the LMH concepts *Greek-alphabet*, *Hebrew-alphabet* and *Roman-alphabet* each serve to differentiate the concept *alphabet*. This is a useful ontological distinction that contributes to the definition of individual letter concepts like *Alpha*, *Beta* and *Gimel*. Since we can represent this specialization pattern via a differentiation set $D_{alphabet}$ as follows:

$D_{alphabet} = \{$*Greek, Hebrew, Roman*$\}$

More generally, the differentiation set of a concept H comprises the set of all concepts M such that the LMH concept M-H is in C*. Thus we have:

$D_{deity} = \{$*Hindu, Roman, Greek,*$\ldots\}$
$D_{architecture} = \{$*Greek, Roman,* $\ldots\}$
$D_{calendar} = \{$*Muslim, Jewish, Hebrew,* $\ldots\}$

We use D to denote the set of all differentiation sets that are implied by $C*$, allowing us to define the absolute affinity between two modifier terms c_1 and c_2 in terms of differentiation as follows:

$$A_{abs}(c_1, c_2) = |\{x \in D : c_1 \in x \wedge c_2 \in x\}| \quad (1)$$

This simply counts the number of base concepts that c_1 and c_2 can both differentiate. We thus define the relative affinity between two modifier terms c_1 and c_2 as follows:

$$A_{rel}(c_1, c_2) = |\{x \in D : c_1 \in x \wedge c_2 \in x\}| \,/\, |\{x \in D : c_1 \in x \wedge c_2 \in x\}| \quad (2)$$

A relative affinity of 1.0 means that both terms differentiate exactly the same concepts in WordNet. It follows that the higher the relative affinity between c_1 and c_2, then the greater the likelihood that a concept differentiated by c_1 can also be differentiated by c_2, while the higher the absolute affinity, the more reliable this likelihood estimate becomes. Affinity thus provides an effective basis for formulating the transformation rules in T*.

We should naturally expect near-synonymous modifiers to have a strong affinity for each other. For instance, *Jewish* and *Hebrew* are near-synonyms because WordNet compounds *Jewish-Calendar* and *Hebrew-Calendar* are themselves synonymous. This is clearly a form of contextual synonymy, since *Jewish* and *Hebrew* do not mean the same thing. Nonetheless, their affinity can be used to generate new compounds that add value to WordNet as synonyms of existing terms, such as *Jewish-alphabet*, *Hebrew-Religion*, and so on.

Recall that literal compounds represent a yoking together of two or more ontological branches. In exploring the space of novel compounds, it will be important to recognize which branches most naturally form the strongest bonds. Another variant of affinity can be formulated for this purpose:

$$A_{domain}(x, y) = |D_x \cap D_y| \quad (3)$$

For instance, A_{domain}(sauce, pizza) = 2, since in WordNet the modifier overlap between the *pizza* and sauce *domains* is {*anchovy, cheese*}.

4 Creative Exploration in the LMH Space

We consider as an exploratory starting point any LMH concept M-H in C*. We can transform this into another concept M-H by replacing M with any M' for which:

$$M' \in \{x | x \in D - \{D_H\} \wedge M \in x\} \quad (4)$$

This formulation may suggest a large range of values of M'. However, these candidates can be sorted by A_{rel}(M, M'), which estimates the probability that a given M'-H will later be validated as useful. One rule in T* can now be formulated for our further consideration:

$$T* : M_1 - H_1 \wedge M_1 - H_2 \wedge M_2 - H_1 \Rightarrow M_2 - H_2 \quad (5)$$

This rule allows the LMH space to be explored via a process of modifier modulation. Suppose we choose *Greek-deity* as a starting point. Since M = *Greek* and H = *Deity*, we can choose M' from any set other than D_{deity} that contains *Greek*:

$D_{alphabet}$ = {*Hebrew, Greek, Roman*}
D_{deity} = {*Greek, Roman, Norse, Hindu, . . .*}

These differentiation patterns suggest that new compounds can meaningfully be created by yoking the ontological branches of *alphabet* and *deity* together. Thus, from $D_{alphabet}$ we can choose M' to be either *Hebrew* or *Roman*, leading to the creation of the LMH concepts *Hebrew-deity* and *Roman-deity*. One of these, *Roman-deity*, already exists in C*, but another, *Hebrew-deity* is novel in a way that Boden terms psychologically or P-Creative [1], inasmuch as it is neither in C_w nor C*. It may thus be of some value as a hypernym for existing WordNet concepts like *Yahwe* and *Jehovah*.

Rule (5) is a general principle for ontological exploration in the space of compound terms. Consider the compound *software-engineering*, which, following (5), is suggested by the joint existence in WordNet of the concepts *software-engineer*, *automotive-engineer* and *automotive-engineering*. While this particular addition could be predicted from the application of simple morphology rules, the point here is that a single exploration principle like (5) can obviate the need for a patchwork of such simple rules.

Of course, one can imagine rules other than (5) to exploit the regularities inherent in WordNet definitions. For instance, consider the sense *gasoline-bomb*, which WordNet glosses as: "a crude incendiary bomb made of a bottle filled with flammable liquid and fitted with a rag wick". By determining which definite description in the gloss conforms to the modifier - in this case it is "flammable liquid" - other modifiers can be found that also match this description. Thus, the new concepts *methanol-bomb* and *butanol-bomb* can be generated, and from this the creative concept *alcohol-bomb* can be generalized. However, each strategy raises its own unique issues, so for now we consider a T* comprising (5) only.

4.1 The Evaluation Mechanism E

For purposes of ascribing value or ontological utility to a new LMH concept M'-H, the concept must first be placed into one of the following categories:

a. M'-H already exists in C* is thus ascribed zero value as an addition to C*.

b. M'-H does not exist in C* but does exist in C_w, and thus corresponds to an existing non-literal concept (such as *monkey-bread*). While it may have value if given a purely literal reading, it cannot be added to C_w without creating ambiguity, and so has zero value.

c. Using R*, M'-H can be seen to describe a non-empty class of existing concepts in C_w, and would thus have value as either a synonym (when this set is a singleton) or as a new organizing super-type (when this set is a severalton). In this case, we say that M'-H has been *internally validated* against C_w.

d. Using a textual analysis of a large corpus such as the World-Wide-Web, M'-H is recognized to have a conventional meaning in C even if it is not described in C_w. In this case, we say that M'-H has been *externally validated* for inclusion in C_w. The fact that M'-H is novel to the system but not to the historical context of the web suggests that it is merely a psychologically or P-Creative invention in the sense of Boden [1].

e. M'-H is recognized to have a hypothetical or metaphoric value within a comprehension framework such as conceptual blending theory (e.g., see Veale et al. [12]), mental space theory, etc. In this case, M'-H may truly be a historically or H-Creative invention in the sense of Boden [1].

In general, a new compound has value if its existence is suggested by, but not recognized by, the lexical ontology. As noted in the introduction, this value can be realized in a variety of ways, e.g., by automatically suggesting new knowledge-base additions to the lexical ontologist, or by providing potentially creative expansions for a user query in an information retrieval system (see [11]).

4.2 Validating New Concepts

The evaluation strategies (c) and (d) above suggest two ways of validating the results of new compound creation: a WordNet-internal approach that uses the structure of the ontology itself to provide evidence for a compound's utility, and a WordNet-external approach that instead looks to an unstructured archive like the web. In both cases, a new compound is validated by assembling a support set of precedent terms that argue for its meaningfulness.

4.3 Internal Validation

The internal support-set for a new compound M-H is the set of all WordNet words w that have: (i) at least one sense that is a hyponym of a sense of H; and (ii) a sense that contains M or some variant of it in its gloss. For instance, the novel compound "rain god" is internally validated by the word set {"Thor", "Parjanya", "Rain giver"}.

When *w* is polysemous, two distinct senses may be used, reflecting the fact that M-H may be metonymic in construction. For instance, the compound "raisin-wine" can be validated internally by the polysemous word "muscatel", since one sense of "muscatel" is a kind of wine, and another, a kind of grape, has a WordNet gloss containing the word "raisin". From this perspective, a "raisin wine" can be a wine made from the same grapes that raisins are made from. Likewise, the compound "Jewish robot" can be validated by simultaneously employing both senses of "Golem" in WordNet, which defines "Golem" as either a Jewish mythical being or as a robotic automaton.

Creative products arise when conceptual ingredients from different domains are effectively *blended* (see Veale and O'Donoghue [12]). It follows that a creative product can be validated by deblending it into its constituent parts and determining whether there is a precedent for combining elements of these types, if not these specific elements. We can thus exploit this notion of deblending to provide internal validation for new compounds. For instance the WordNet gloss for pizza lists "tomato sauce" as an ingredient. This suggests we can meaningfully understand a compound of the form M-*pizza* if there exists a compound M-*sauce* that can be viewed as a replacement for this ingredient. Generalizing from this, we can consider a new compound M_1-H_1 to be internally validated if H has a sense whose gloss contains the compound M_2-H_2, and if the ontology additionally contains the concept M_1-H_2. It follows then that the novel compounds *apple-pizza, chocolate-pizza, taco-pizza*, and *curry-pizza* will all be internally validated as meaningful (if not necessarily enjoyable) varieties of pizza.

4.4 External Validation

In contrast, the external validation set for a compound M-H is the set of distinct documents that contain the compound term "M H", as acquired using a web search engine. For instance, given the WordNet concepts *naval-engineer*, *software-engineer* and *naval-academy*, rule (5) generates the hypothesis *software-academy*, which cannot be validated internally yet which retrieves over 1000 web documents to atest to its validity.

This web strategy is motivated by Keller and Lapata's [3] finding that the number of documents containing a novel compound reliably predicts the human plausibility scores for the compound.

Nonetheless, external validation in this way is by no means a robust process. Since web documents are not sense tagged, one cannot be sure that a compound occurs with the sense that it is hypothesized to have. Indeed, it may not even occur as a compound at all, but as a coincidental juxtaposition of terms from different phrases or sentences. Finally, even if found with the correct syntactic and semantic form, one cannot be sure that the usage is not that of a non-native, second language learner. These possibilities can be diminished by seeking a large enough sample set, but this has the effect of setting the evidential bar too high for truly creative compounds. However, another solution lies in the way that the results of external validation are actually used, as we shall later see.

4.5 Validating New Synonyms

Many of the compounds that are validated either by internal or external means will be synonyms of existing WordNet terms. As such, their creative value will not represent an innovative combination of ideas, but rather a creative use of paraphrasing. The nature of (5) makes it straightforward to determine which is the case.

In general, when M_1-H_1 and M_2-H_1 are themselves synonyms, then M_2-H_2 will be a synonym of M_1-H2. For instance, from the combination of *applied-science, engineering-science* and *applied-mathematics*, we can generate from (5) the new compound *engineering-mathematics*. This compound cannot be validated internally, but since it retrieves more than 300,000 documents from the web, this is enough to adequately atest to its meaningfulness. Now, since *applied-science* and *engineering-science* are synonymous in WordNet, we can conclude that *engineering-mathematics* and *applied-mathematics* are themselves synonymous also.

4.6 Creativity in the Validation Gap

The difference between internal and external validation strategies can be illuminating. Internal validation verifies a compound on the basis that it *could meaningfully* exist, while external validation verifies it on the basis that it *does actually* exist in a large corpus. Therefore, if a compound can be validated externally but not internally, it suggests that the concept may by P-Creative. In contrast, if the compound can be validated internally but not externally, it suggests that the compound may be H-Creative and represent a genuine historical innovation (if only a lexical one, and of minor proportions).

For instance, the new compound "sea dance" (analogous to "rain dance") cannot be validated internally, yet can be found in over 700 internet documents.

It thus denotes a P-Creative concept. In contrast, the compound "cranial vein" yields no documents from a web query (on AltaVista), yet can be internally validated by WordNet via the word-concept *Diploic-Vein*, a blood vessel that serves the soft tissue of the cranial bones. Likewise, the compounds "chocolate pizza", "taco pizza" and many more from the yoking of D_{pizza} and D_{sauce} can all be validated externally via hundreds of different web occurrences, and so represent P-Creative varieties of pizza. However, compounds like "Newburg pizza" (a pizza made with lobster sauce) and "wine pizza" (a pizza made with wine sauce) can only be validated internally and are thus candidates for H-Creative innovation.

5 Large-Scale Evaluation

A large scale evaluation of these ideas was conducted by exhaustively applying the T* rule of (5) to the noun taxonomy of WordNet 1.7. To better see the effect of affinity between modifiers, Table 1 ranks the results according to the measure A_{rel} from (1).

A_{rel}	1	2	3
No. of compounds generated	941,841	22,727	2,175
% H-Creative	0.49%	0.63%	1.38%
% P-Creative	35.65%	33.77%	34.57%
% Conflations	0.10%	0.10%	0.05%
%Indeterminate	63.76%	65.49%	64.00%

Table 1. Number of compounds created, and their assessment, for each affinity level.

Conflations are terms that exist both as compounds and as conflated lexical atoms. For instance, while the compound "bull dog" may not exist in WordNet, its conflation "bulldog" does. Compound discovery is thus a useful means of re-expanding these conflations when it is meaningful to do so.

As one might expect, lower affinity levels allow greater numbers of new compounds to be created. Interestingly, however, Table 1 suggests that as the affinity threshold is raised and the number of compounds lowered, the creativity of these compounds increases, as measured by the relative proportion of H-Creative terms that are generated.

Generating compound terms in a lexical ontology is a creative process that demands rigorous validation if the ontology is not to be corrupted. Of the two strategies discussed here, external validation is undoubtedly the weaker of the two, as one should be loathe to add new compounds to WordNet on the basis of web evidence alone. However, external validation does serve to illustrate the soundness of internal validation, since 99.51% of internally validated concepts (at $A_{rel} = 1$) are shown to exist on the web. It follows then that the *absence* of external validation yields a very conservative basis for assessing H-Creativity. Web validation is perhaps better used therefore as a means of rejecting creative products than as a means of discovering them. In fact, when used as a reverse barometer in this way, the inevitable errors that arise from web-based validation serve only to make the creative process more selective.

6 Conlusions and Future Work

We are currently considering ways of broadening the scope of internal validation while maintaining its conceptual rigour. This should counter-balance the high rejection rate caused by an overly conservative external validation process, and thereby allow us to identify a higher percentage of H-creative products. As shown with the "pizza" examples of section 4.3, we have already begun to explore the possibilities of validation latent in the WordNet ontology itself. So while the use of web content for external validation suggests that creative discovery has a role to play in producing and expanding web queries, internal validation remains our central thrust, leading to what we hope will be a new, more creative model of the thesaurus.

In grounding our discussion in the creative framework of Boden [1] and its formalization by Wiggins [14], we have placed particular emphasis on the labels P-Creative and H-Creative. However, the empirical results of section 5 suggest that this binary categorization may be overly reductive, and that a more gradated system of labels is needed. For instance, the novel compounds *computer-consultant* and *handwriting-consultant* are both created from a yoking of the domains *expert* and *consultant*, and because each is externally validated, each is considered P-Creative. However, while only a handful of documents can be marshalled to support *handwriting-consultant*, the amount of web evidence available to support *computer-consultant* is vast. So it is wrongheaded to consider both as equally P-Creative and lacking in H-Creativity, since the dearth of existing uses suggests *handwriting-consultant* has far greater novelty. Perhaps what is needed then is not a binary categorization but a continuous one, a numeric scale with P- and H-Creativity as its poles. This scale would function much like the continuum used by [7] to separate banal metaphors (which he dubbed *epiphors*) from creative ones (or *diaphors*).

References

1. Boden, M.A. 1990. The creative mind: Myths and mechanisms, Basic Books, New York.
2. Hanks, P. 2004. WordNet: What is to be done? In the proceedings of GWC'2004, the 2nd Global WordNet conference, Masaryk University, Brno.

3. Keller, F. and Lapata, M. 2003. Using the web to obtain frequencies for unseen bigrams. Computational Linguistics.
4. Lapata, M. & Lascarides,A. 2003. Detecting Novel Compounds: The Role of Distributional Evidence. In Proceedings of the 11th Conference of the European Chapter for the Association of Computational Linguistics, 235-242. Budapest.
5. Lin, D. 1998. Extracting Collocations from Text Corpora. Workshop on Computational Terminology. Montreal, Canada
6. Lynott, D. and Keane, M. 2003. The role of knowledge support in creating noun-noun compounds. In the proceedings of the 25th Conference of the Cognitive Science Society.
7. MacCormac, E. R. 1985. A Cognitive Theory of Metaphor. Cambridge, MA: MIT Press.
8. Miller, G., Beckwith, R., Fellbaum, C., Gross, D. and Miller. K.J. 1990. Introduction to WordNet: an on-line lexical database. International Journal of Lexicography, 3(4), 35 - 244.
9. Sag, I.A., Baldwin, T., Bond, F., Copestake, A. and Flickinger, D. 2002. Multiword Expressions: A Pain in the Neck for NLP, In Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002), Mexico City.
10. Segond, F. & Tapanainen, P. 1995. Using a Finite-State Based Formalism to Identify and Generate Multiword Expressions. MLTT technical Reports. Ref.: 1995-019
11. Veale, T. 2004. Creative Information Retrieval. In the proceedings of CICLing, A. Gelbukh, ed. LNCS 2945, Springer: Berlin.
12. Veale, T. and O'Donoghue, D. 2000. Computation and Blending. Cognitive Linguistics, 11(3/4): 253-281.
13. Vendler, Z. 1967. Linguistics and Philosophy. Ithaca, New York: Cornell University Press.
14. Wiggins, G. 2003. Categorizing Creative Systems. In the proceedings of the 3rd Workshop on Creative Systems, IJCAI'03, Acapulco, Mexico.

Tapping the Entrenched Creativity of Lexical Ontologies

Tony Veale

Dept. Of Computer Science
University College Dublin
Belfield Dublin 4
Ireland
tony.veale@ucd.ie

Abstract. Creativity is an intriguing cognitive phenomenon that we encounter through our interactions both with the world and with the words we choose to describe the world. As such, linguistic creativity opens a window into more general creative processes that, over time, become lexically entrenched in a language. This entrenchment can take a variety of forms, but one of the richest veins to tap is that of polysemy, the phenomenon whereby a word can assume multiple related meanings. In this paper we describe techniques for identifying those instances of polysemy in a lexical ontology like WordNet that are most conducive to a creative reasoning system. We argue that generalizations based on these instances allow a system to be not only more linguistically creative, but more creative in a non-linguistic manner as well. Along the way we consider and resolve one of the paradoxes that arises when trying to generate creative products from an ontology, lexical or otherwise: the need to generate innovative ideas that are at once both substantially different from what already exists and at the same time understandable in terms of existing ideas.

1 Introduction

Creativity is a vexing phenomenon to pin down formally [1], which is perhaps why we tend to think of it in largely metaphoric terms. For example, creativity is often conceived as a form of mental agility that allows gifted individuals to make astonishing mental leaps from one concept to another [2]. Alternately, it is popularly conceived as a form of lateral thinking that allows those who use it to insightfully cut sideways through the hierarchical rigidity of conventional categories [3]. Common to most of these metaphors is the idea that creativity involves recategorization, the ability to meaningfully move a concept from one category to another in a way that unlocks hidden value, perhaps by revealing a new and useful functional property of the concept. For example, psychometric tests such as the Torrance test of creative thinking [4] try to measure this ability with tasks that, e.g., ask a subject to list as many unusual and interesting uses of old tin cans as possible.

The ad-hoc nature of creativity is such that most ontologies, perhaps all ontologies, do not and can not provide the kinds of lateral linkages between concepts to allow this kind of inventive recategorization. Instead, ontologies tend to concentrate their representational energies on the hierarchical structures that, from the lateral thinking perspective, are as much a hindrance as an inducement to creativity. This is certainly true of WordNet [5], whose *isa* hierarchy is the most richly developed part of its lexical ontology, but it is also true of language-independent ontologies like Cyc [6], which are rich in non-hierarchical relations but not of the kind that capture deep similarity between superficially different concepts. Unfortunately, it is precisely connections like these that most readily fuel the recategorization process.

However, because WordNet is an ontology of lexicalized concepts, it necessarily captures much of the lexical creativity evident in everyday language. Often, this word-use is a reflection of deeper recategorization processes at the conceptual level. In this paper we present techniques for identifying and extracting this evidence using automatic and semi-automatic means. Once tapped, these entrenched structures serve as a basis for augmenting WordNet with the additional category-broadening connections between concepts that are necessary to facilitate the creative processes of dynamic reconceptualization.

2 Structure and Function in Lexical Ontologies

When choosing the appropriate ontological location at which to situate a given concept, WordNet's designers typically choose just a single location[1]. A key concern then in using WordNet and similar lexical ontologies for the generation of creative products is the way in which these ontologies, through an overuse of single-inheritance taxonomic organization, conflate the different dimensions of structure, behaviour and function [13]. For instance, an ontologist, if forced to choose, might taxonomize a receptacle under the super-type *{hollow_object}* if seeking a structural perspective, or under the super-type *{container}* if seeking a functional perspective. Naturally, this has serious ramifications for models of semantic similarity that exploit taxonomic structure (see [11] for a review), since concepts that are very similar in one dimension (e.g., structure) may be very different in another (e.g., function), and may thus be located in very different parts of the ontology. In such ontologies then, it is not clear what exactly is being measured by a semantic similarity metric, or even whether similarity scores for different concept pairings are directly comparable. The ramifications for creative processing within an ontology are therefore equally serious. The most creative containers, such as hollow tree trunks, dental cavities, etc., will not be taxonomized as such, since *{container}* is a teleological perspective that does not apply to natural kinds. What is needed is a way for a creative system to determine that *{container}* and *{hollow_object}* are fundamentally compatible perspectives. In a multiple-inheritance ontology, we could expect most receptacle concepts to explicitly specialize both *{container}* and *{hollow_object}*, allowing a creative system to at least determine a statistical correlation between these categories. In a single-inheritance lexical ontology, or one where multiple-inheritance is rarely used, a system must rely on other means.

---

[1] WordNet supports the use of multiple-inheritance, but it's usage is rare indeed. However, since the semantics of multiple inheritance are unclear even for artificial programming languages, this is perhaps not so surprising.

Because an eschewal of multiple-inheritance forces an ontologist to also eschew a holistic view of each concept, such ontologies can suffer from extreme *conceptual fragmentation*. This is particularly true of lexical ontologies like WordNet, where an ontologist/lexicographer has the freedom to split a complex word-concept into a range of different splinter senses that belong in different taxonomic locations. For instance, to shoehorn a rich concept like *disaster* into WordNet, its designers first splinter it into an *{event}, {act}* and *{state}* perspective; this tripartite form fits the word-usage data but does not do representational justice to the concept itself. Likewise, the word *knife* is splintered into separate *{edge-tool}* and *{weapon}* senses. But this is surely a false discrimination to make, since a knife is a weapon precisely because it possesses a sharp edge, just as many receptacles derive their functionality as containers from their hollow structure. So by splintering the whole into smaller senses, an ontology loses the important relationship between structure and function that one must necessarily understand before the concept can be used creatively.

Lexicographers refer to this splintering as *polysemy*, a form of lexical ambiguity in which a word has multiple related meanings. Normally these meanings arise from the systematic use of metonymy, wherein a word comes to denote both its original meaning and a related contiguous meaning (e.g., over time, words for containers also come to denote the contents of those containers, allowing one to "boil the kettle" or "drink a bottle"). The form of polysemy that interests us most from a creativity perspective is that which is function-transforming yet structure-preserving. Instances of this kind of polysemy reveal, in entrenched lexical terms, the way certain objects have in the past been extended to serve new functions. For instance, English has a variety of words that denote both animals and the meat derived from them (e.g., *chicken, lamb, cod*), and this polysemy reflects the transformation potential of animals to be used as meat. Likewise, the polysemy of *knife* reveals the potential for an object with a sharp edge (structure preservation) to be used as a weapon (function transformation).

Thus, if we can identify all such instances of function-transforming yet structure-preserving polysemy in WordNet, we can generalize from these a collection of pathways that allow a system to hypothesize creative uses for other concepts that are not so entrenched via polysemy. Consider again the different but related WordNet senses of *knife*: one is an *{edge-tool}* used for cutting, and one is a *{weapon}* used for injuring. Each sense describes structurally similar objects (sharp flat objects with handles) with a common behavior (cutting) that differ primarily in function (i.e., slicing vs. stabbing). This polysemy suggests a generalization that captures the functional potential of any other *{edge-tool}*, such as *{scissors}* and *{shears}*, to also be used as a *{weapon}*.

More formally, consider a polysemous word ω with a pair of related senses $<\omega_1, \omega_2>$. Suppose the hypernym h_1 of ω_1 expresses a functional perspective $fun(\omega_1)$, while the hypernym h_2 of ω_2 expresses a structural perspective $structure(\omega_2)$. We can generalize this relationship by assuming that other objects with the same structural properties can also serve the same functional role. That is,

$$h_2(x) \rightarrow h_1(x)$$

This generalization might allow us to infer that all hollow objects can be used as containers, or that all sharp-edged tools can be used as weapons. The problem is, however, that we cannot know, at least using automated means, that h_1 expresses a functional perspective and that h_2 expresses a structural perspective, since WordNet is not annotated with such information, either explicitly in its structure or implicitly in its glosses.

We can, alternately, capture a similar generalization between h_1 and h_2 if we know that $h_1$is a broader category that h_2. Broadness is a relatively straightforward aspect of category structure to measure in ontological terms, since broader categories reside in higher levels of the ontology. In terms of a lexical ontology like WordNet, we can define broadness as a relative measure of the number of ways in which a concept can be specialized. Thus, we consider h_1 to be broader than h_2 if h_1 has more descendent hyponyms than h_2. Since by this metric *{weapon}* is a broader category than *{edge-tool},* we can infer that other edge-tools such as axes and scythes may be used as weapons too. Conversely, we do not infer that all weapons are potential edge-tools. In effect, the generalization represents an inductive hypothesis that, while never fully articulated, captures the idea that it is the sharp edge of a tool that allows it to be used as a weapon.

3 Identifying Creativity-Supporting Polysemy in WordNet

Before we can tackle the problem of recognizing particular kinds of function-transforming polysemy in WordNet, we must first resolve the more basic issue of recognizing polysemy at all. This is a very considerable problem since not all instances of lexical ambiguity are instances of polysemy, and WordNet fails to separate those senses which are related by meaning or etymology from those that merely share the same lexeme for reasons of historical coincidence.

The research literature provides a number of automated approaches to recognizing truly polysemous relationships, which we may divide into top-down and bottom-up categories. A paradigmatic example of the top down category is the use of hand-crafted lexical rules to capture broad regularities in the workings of polysemy. In WordNet, for instance, cousin relations [5,7] are manually established between concepts in the upper-ontology to explain the systematicity of polysemy at lower levels. Thus, once a connection between *{animal}* and *{food}* is established, it can be instantiated by words with both an animal and food sense, suggesting that these words exhibit an *animal:food* polysemy. However, this approach is limited by the number of high-level connections that are manually added, and by the need to list often copious exceptions to the pattern (e.g., *mate* the animal partner, and *mate* the berry drink, are merely homonyms; the latter is not derived from the former).

Conversely, a paradigmatic example of a bottom-up approach is the use of statistically-observed distributional patterns to infer systematic behavior among words [8, 9]. In this approach, families of words with similar sense breakdowns are first recognized in the lower ontology, and then generalized to yield connections between higher-level concepts [8, 9, 10]. For instance, many words have senses that denote both a kind of music and a kind of dance (e.g., *waltz, tango, conga*), which suggests a polysemous relationship between *{music}* and *{dance}*. Likewise, over 140 different words in WordNet have both a *{person}* and a *{language}* sense, strongly suggesting the presence of a *speaker:spoken* metonymy at work below the surface.

Both of these approaches treat polysemy as a systematic phenomenon best described at the level of word families. However, while such a treatment reveals interesting macro-tendencies in the lexicon, it does little to dispel the possibility that homonymy might still operate on the micro-level of individual words (as demonstrated by the size of the exception list needed for the cousins approach [5,7]). It also fails to recognize patterns that are cognitively significant but statistically under-represented in the lexicon. For instance, only one entry in WordNet – *florist* – has both a *{businessperson}* and a *{place_of_business}* sense, yet the metonymic relationship between both is one that is commonly observed in everyday language and thought. We thus prefer to use an evidential case-by-case approach to detecting polysemy, connecting a pair of senses only when explicit local taxonomic evidence can be found to motivate a connection. This evidence can take many forms, so a patchwork of different strategies is required. These strategies do not utilize distributional information, and so are as adept at recognizing ad-hoc metonymies as those that are significantly more entrenched. We describe now the three most interesting of these strategies.

The coverage of each heuristic strategy is estimated relative to that achieved by the *cousins* collection of 105 regular polysemy noun-sense groupings that are hand-coded in WordNet [7]. Over-generation is estimated relative to the overlap with the cousins exception list [7], which permits us to also estimate the accuracy of each heuristic.

Explicit Ontological Bridging: a sense pair $<\omega_1, \omega_2>$ for a word ω can be linked if ω_1 has a hypernym that can be lexicalized as M-H and ω_2 has a hypernym that can be lexicalized as M, the rationale being that ω_2 is the M of ω_1 and ω_1 is the H of ω_2. E.g., the word *olive* has a sense with a hypernym *{fruit-tree}*, and another with the hypernym *{fruit}*, therefore M = *fruit* and H = *tree*. *(Coverage: 12%, Accuracy: 94%).*

Hierarchical Reinforcement: if $<\alpha_1, \alpha_2>$ and $<\beta_1, \beta_2>$ are sense pairs for two words α and β where α_1 is a hypernym of β_1 and α_2 is a hypernym of β_2, then $<\alpha_1, \alpha_2>$ reinforces the belief that $<\beta_1, \beta_2>$ is polysemous, and vice versa. For example, *herb* denotes both a plant and a foodstuff in WordNet, and each of these senses has a hyponym that can be lexicalized as *sage*. *(Coverage: 7%, Accuracy: 12%).*

Cross-Reference: if <ω_1, ω_2> is a sense pair for a word ω and the WordNet gloss for ω_2 explicitly mentions a hypernym of ω_1, then ω_2 can be seen as a conceptual extension of ω_1. For instance, the railway-compartment sense of *diner* mentions *restaurant* in its gloss, while another sense actually specifies *{restaurant}* as a hypernym. This suggests that the railway sense is an extension of the restaurant sense that uses the later as a ground for its definition. *(Coverage: 62%, Accuracy: 85%)*.

3.1 Function-Transforming Structure-Preserving Polysemy

These heuristic strategies are very effective at arguing for polysemy on the local merits of individual words. However, for every creatively-useful instance of polysemy like *knife* (*{weapon}* versus *{edge-tool}*), we encounter a truly unhelpful instance like *capsule* (*{space-vehicle}* versus *{medicine}*). One cannot meaningfully reuse aspirin-capsules as spacecraft, or vice versa, except in a humorous context. This is not to say that such examples are degenerate cases of polysemy, since one can readily intuit the rationale for using the word *capsule* in such different contexts. Rather, it is the specific purpose to which we stretch each instance of polysemy, namely creative reuse, that makes these examples unacceptable.

It follows that only those instances of polysemy that relate concepts with compatible structural forms can be reliably used for creative purposes. By compatible structural form we mean more than a general physical resemblance; after all, space capsules and gel-capsules do share certain spatial characteristics like shape, and it for this reason that they merit the same lexical label *capsule*. More important to structural compatibility is the dimension of scale, for if this dimension is not respected our creative reasoner is likely to suggest creative reuses that are novel by virtue of insensibility and practical only as an exercise in the surreal. Of course, this absurdity may well be creative in a humorous context, for humour is as valid a medium for human creativity as any other. Nonetheless, practical creative products can only arise from reuses that respect physical constraints of the objects concerned.

Having conceded this limitation, we are faced with a dilemma: while WordNet contains enough knowledge to facilitate the automated detection of polysemy, it lacks sufficient knowledge to allow an automated detection of structure-preserving polysemy. Lacking a reliable way to automate this task, we resort to supervised annotation and manually filter those instances of polysemy that are not structure-preserving. Fortunately, this task is made less onerous if, instead of filtering the instances of polysemy themselves, we instead filter the generalizations made from these instances. This filtering, which can be done in a matter of hours, causes almost 50% of polysemous sense pairings to be rejected.

4 Types of Ontological Creativity

The polysemy relationships that can be extracted from WordNet are merely the residue of past creativity by the language community. However, new creative insights can be generated by generalizing from these entrenched precedents, to either broaden existing categories and admit new members not previously considered eligible, or to re-categorize members of existing categories under different branches of the ontology.

The polysemy-based generalizations described in sections 2 and 3 provide an interesting means of realizing the processes of recategorization and category-broadening in a lexical ontology. Each such generalization can, in a licensing context, be considered a special kind of *isa* relationship. Processes that normally exploit *isa* relationships in an ontology will now encounter a transformed search-space in which semantically-distant categories that were not previously reachable are now but a link or two away.

4.1 Category Broadening

Imagine we want to broaden the WordNet category *{weapon}*. The members of this category can be enumerated by recursively visiting every hyponym of the category, which will include *{knife}, {gun}, {artillery}, {pike}*, etc. However, by licensing the use of polysemy-based generalizations in addition to the standard *isa* links of the ontology, additional prospective members can be reached and admitted on the basis of their functional potential. Thus, the polysemy of *knife* causes not only *{dagger}* and *{bayonet}* but *{steak_knife}* and *{scalpel}* to be visited. Stretching category boundaries even further, the generalization *edge_tool(x)* $\rightarrow$ *weapon(x)* allows the category *{edge_tool}* to be subsumed in its entirety, thereby allowing *{scissors}, {axe, ax}, {razor}* and all other sharp-edged tools to be recognized as having weapon-like potential.

Category broadening is a very revealing process, not only about the functional potential of everyday objects, but also about the inevitable gaps in an ontology like WordNet. For instance, the category *{apparel, clothing, clothes}* can be broadened to admit baseball gloves, anklets, metal helmets, furs and animal skins, while the category *{medicine, medication}* can be broadened to admit toiletries and oleoresins, and the category *{food}* can be broadened to admit a variety of potentially edible substances, some too disgusting to list here.

4.2 Category Hopping

Imagine, following the Torrance test, we want to move the concept *{coffee_can}* to a new category that will offer a functional perspective on how to effectively reuse old tin cans. The existing WordNet categories that house *{coffee_can}* can be enumerated by recursively visiting each of its hypernyms in turn, which will include *{can, tin_can}*,*{container}* and *{artifact}*. Now, each of these hypernyms is a potential point of departure to another category if, as well as traversing *isa* relations, we use polysemy-based generalizations to slip from one rail of the ontology to another. WordNet defines *{coffee_can}* as a hyponym of *{can, tin_can}*, and from here a leap can be made to *{steel_drum, drum}*, since both are hyponyms of *{container}* whose glosses further specify them as kinds of *metal container*. From *{steel_drum, drum}* there exists a polysemy link to *{tympan, membranophone, drum}*, a non-container artifact which WordNet defines as a hyponym of *{percussion_instrument}*. This chain of reasoning, from *{coffee_can}* to *{tin_can}* to *{steel_drum}* to *{tympan, membranophone, drum}*, supports the creative insight that allows an old tin can to be used a musical drum. Central to this insight is the polysemy of *drum* and the generalizations that can be derived from it.

In general, polysemy supports creativity by providing just one very important link in the recategorization chain. A *{dog collar}* can be fashionably reused as a *{necklace}* because the polysemy of *collar* links *{collar}* to *{choker, collar}*. We can meaningfully think of jewelry as a piece of fine-art (and thus consider exhibiting it in a gallery) because of the polysemy of *gem* that links *{gem, jewel}* to *{gem, treasure}*. Likewise, we can think of photography as a fine art because photograph and art collide via the polysemy of *mosaic*, *vignette* and *scene*.

5 Creativity, Utility and Similarity

Some recategorizations will exhibit more creativity than others, largely because they represent more of a mental leap within the ontology. We can measure this distance using any of a variety of taxonomic metrics [11], and thus rank the creative outputs of our system. For instance, it is more creative to reuse a coffee can as a *{percussion_instrument}* than as a *{chamberpot, potty}*, since like *{tin_can}* the latter is already taxonomized in WordNet as a *{container}*. Any similarity metric (called σ) that measures the relative distance to the lowest common hypernym will thus attribute greater similarity to *{coffee_can}* and *{potty, chamberpot}* than to *{coffee_can}* and *{tympan, drum, membranophone}*. This allows us to measure the creative distance in a recategorization from α to γ as $1 - \sigma(\alpha, \gamma)$.

Of course, distance is not the only component of creativity, as any recategorization must also possess some utility to make it worthwhile (e.g., there is a greater distance still between tin cans and fish gills, but the former cannot be sensibly reused as the latter). In other words, a creative product must be unfamiliar enough to be innovative but familiar enough to be judged relative to what we know already works. This is the paradox at the heart of ontological creativity: to be creative a recategorization must involve a significant mental leap in *function* but not in *form*, yet typically (e.g., in WordNet), both of these qualities are ontologically expressed in the same way, via taxonomic structure. This suggests that taxonomic similarity σ

must be simultaneously maximized (to preserve structural compatibility) and minimized (to yield a creative leap).

Fortunately, polysemy offers a way to resolve this paradox. If a creative leap from α to γ is facilitated by a polysemous link from $<\beta, \gamma>$, the sensibility of the leap can be measured as $\sigma(\alpha, \beta)$ while the creativity of the leap can be measured as $1 - \sigma(\alpha, \gamma)$. The value of a creative product will be a function of both distance and sensibility, as the former without the latter is unusable, and the latter without the former is banal. The harmonic mean is one way of balancing this dependency on both measures:

$$value(\alpha, \gamma) = 2\sigma(\alpha, \beta)\,(1 - \sigma(\alpha, \gamma)) \;/\; (1 + \sigma(\alpha, \beta) - \sigma(\alpha, \gamma))$$

Other variations on this formula can be used to give greater of lesser weight to the roles of sensibility and distance in determining the value of a creative insight.

6 Concluding Observations

As noted earlier, WordNet lacks the key knowledge structures and annotations that would directly facilitate creative reasoning. In this paper we have endeavored to show that some of this knowledge is actually implicit in the polysemous structure of the lexicon, so that once tapped, some productive creative reasoning can be supported. Future directions for this work are many and varied. At present, we feel it would surely be worth the investment in man-years to annotate WordNet with the structure-behavior-function distinctions that are so valuable to creativity research.

The ideas in this paper have now been implemented in a computational system called *Kalos* (a Greek word connoting beauty through fitness of purpose [3]). A collection of 25 different polysemy detectors (of which 3 were described here) achieve 96% of the coverage offered by WordNet’s own cousin relations, at a precision of 85%. In our pilot study, we focused on the subset of these polysemous relations that connect artifactual noun senses, where this subset is hand-filtered to yield 991 instances of behaviour-preserving, function transforming polysemy. Generalizing from these instances and performing a second phase of hand-checking to filter out spurious hypotheses, we are left with 454 inter-category subsumption hypotheses. These generalizations are a powerful addition to WordNet’s upper and middle ontologies, facilitating a creative flexibility in determining category membership that is useful to a variety of applications, from creative writing tools to text understanding systems.

References

1. Wiggins, G. Categorizing Creative Systems, in: Proc of the 3rd Workshop on Creative Systems, IJCAI’03, Acapulco, Mexico. (2003)
2. Hutton, J.: Aristotle's Poetics. Norton, New York (1982)

3. de Bono, E. Parallel Thinking. Viking Press: London (1994)
4. Torrance, E. P. The Torrance Tests of Creative Thinking. Scholastic Testing Service. Bensonville, Illinois. (1990)
5. Miller, G. A.: WordNet: A Lexical Database for English. Communications of the ACM, Vol. 38 No. 11 (1995)
6. Lenat, D., Guha, R.V.: Building Large Knowledge-Based Systems. Addison Wesley (1990)
7. WordNet documentation. www.princeton.edu/~wn/ (2003)
8. Peters, W., Peters, I., Vossen, P. Automatic sense clustering in EuroWordNet. In: Proc of the 1st international conference on Language Resources and Evaluation. Spain (1998)
9. Peters, I., Peters, P. Extracting Regular Polysemy Patterns in WordNet. Technical Report, University of Sheffield, UK. (2000)
10. Peters, W., Peters, I. Lexicalized Systematic Polysemy in WordNet. In the proceedings of the 2nd international conference on Language Resources and Evaluation. Athens. (2000)
11. Budanitsky, A., Hirst, G. Semantic Distance in WordNet: An experimental, application-oriented evaluation of five measures. In: Proc of the Workshop on WordNet and Other Lexical Resources, North-American chapter of ACL. Pittsburgh. (2001)
12. Lakoff, G.: Women, Fire and Dangerous Things. Uni. of Chicago Press: Chicago (1987)
13. Gero, J. S.: Design prototypes: a knowledge representation schema for design, *AI Magazine* **11**(4): 26-36 (1990)

Automating the Interpretation of Novel, Noun-Noun Compounds Using WordNet

Georgios Tagalakis and Mark T. Keane

Department of Computer Science, University College Dublin, Ireland

Abstract. Natural languages are extended by combining existing words into novel combinations to convey new ideas. Many such novel combinations are syntactically characterized as noun-noun compounds. People seem to have well developed abilities for comprehending such novel compounds, even though they often invite multiple, ambiguous meanings. In this paper, we present CoBlLex - a system that comprehends such compounds using ideas from cognitive models of compounds comprehension and standardized knowledge-bases, like WordNet and CoreLex. The resulting system combines the richness of human interpretation patterns with the beneficial property of using an independently-crafted knowledge base. The main components and methods of this system are reported, as is an evaluation of its outputs. The sources of the system's innovation and how creativity is reflected by the proposed methods are also discussed along with future directions of this work.

1 Introduction

On a daily basis, people manifest a type of "mundane creativity" in the ways in which they manipulate words to extend natural languages. In English, most of the new phrases entering the language are combinations of existing words. Many of these new phrases are noun-noun compounds like *blood diamonds*, *helicopter parent*, *information pollution*, *shark politician* and *voodoo programming*. A lot is known about the process of interpreting noun-noun compounds from many studies in Linguistics (e.g, [8, 15]) and Psychology (e.g., [27, 29]). More recently, the understanding of such compounds has attracted the attention of Cognitive and Computer Scientists as they attempt to automate the interpretation process (e.g., [3, 5, 9, 14, 16]).

The present paper is an outgrowth of this work in proposing a program for Conceptual Blending of noun Lexes (CoBlLex) that interprets compounds in a psychologically plausible fashion. As we shall see in the following review, previous related attempts to automate the creativity in compound understanding have not been as successful as they could be.

In the next section we review the relevant computational models of conceptual combination (Section 2) before presenting the CoBlLex algorithmic design and operation (Section 3). In Section 4 we evaluate the system's performance and discuss aspects of creativity as they are reflected in the interpretations the system produces. We conclude with a summary of some of the shortcomings of the current program and recommendations for future work (Section 5).

2 A Review of Relevant Work

While the computational literature on the interpretation of noun-noun compounds is quite extensive, only a few of these programs manage to capture the breadth of human interpretation abilities using reasonably adequate knowledge bases.

Leonard's [14] program was one of the first models to be based on computational analysis of a large corpus of noun-noun sequences (about 2000) drawn from English novels. After forming a listing of the constituent nouns from these extracted compounds, each noun was marked with semantic features, including related overt and covert verbs. A dictionary formed from this knowledge engineering step was then used to identify eight main patterns of interpretation for noun-noun sequences (e.g., locative patterns: *deck chair*, annex: *blood pressure*, material: *paper bundle*, etc.). Leonard's program interpreted a presented noun-noun compound by relating it to one of these patterns and forming an interpretation from this pattern. Leonard argued that 76% of the interpretations generated by the program corresponded to acceptable meanings for English compounds. Similar approaches using taxonomies of interpretations have been proposed in Linguistics and Psychology (e.g., [7, 15, 27]). The main objection to such approaches is that many interpretations can be found that have to be forced into the taxonomy or fall outside of it (cf., [6]). This fact is a major drawback.

Costello & Keane's [5] constraint model aims to provide a first-principles basis for understanding novel, noun-noun compounds. Taking predicate descriptions for constituent concepts of the compound from its knowledge base, it generates a large number of possible meanings for the compound (based on unified subsets of these predicates). This set of possible meanings is then filtered using the three constraints of diagnosticity, plausibility, and informativeness. While this model can generate a very broad, psychologically plausible set of meanings for a given compound it is too prolix, performing too-much first-principles computation. As such, it generates several thousand interpretations for a single compound and does not exploit familiarity patterns of interpretation at all. Finally, the constraint model uses a handcrafted knowledge base. Though this knowledge base was constructed blind to the compounds that were tested using it, the program would benefit, from a generality perspective, if it were able to operate with more standardized knowledge bases.

The PUNC model [17] was designed to overcome the prolixity of the constraint model by being more efficient and sensitive to familiar patterns of interpretation. This program incorporates the same constraints as the previous model. A comparative analysis of the interpretations produced by PUNC with those produced by people yields a strong correlation between PUNC's goodness score and the frequency of production of people's interpretations. Having said this, PUNC still suffers from its reliance on a handcrafted knowledge base. The handcrafting of knowledge in this way always opens the door to criticisms of bias or tailored representations.

We know of at least two models that have tried to use the independent WordNet knowledge base. These are the models of [19] and [12]. McLoughlin [19] examined the hypothesis that novel noun-noun compounds could be interpreted from a training set of compounds and the position of the constituent nouns in a hierarchically-structured lexicon (WordNet, [21]). He compiled a sample data set of about 900 novel compounds from the SUSANNE corpus [25], filtered them out, and allocated the selected ones to 24 interpretation categories. Using these compounds, key links in WordNet are nominated and assigned to these categories using inductive learning strategies. Then the interpretation of the key links are taken to predict interpretations of novel, unseen compounds governed by these links using a cluster strategy.

Unfortunately, McLoughlin's experiments revealed that only 44% of the predictions made by the cluster strategy were correct and that a huge number of training examples would be required to significantly improve performance. An alternative technique (proximity valuation method) that used superordinate links of the target compound, yielded better but still not stellar results (55% correct predictions). Overall, these techniques seem to fail because of misclassifications during data preparation and difficulties encountered with the lexicon (e.g., lack of expected taxonomic relations).

Hayes et. al [12] adopt an approach to concept combination that grows from the work of [8, 15, 27], according to which noun-noun compound can be interpreted in terms of a core set of relations such as *made of*, *location*, etc. Based on this assumption they created a model that interfaces with both WordNet and the Web to create a large-scale model of noun-noun compound interpretation. It is not clear, though, how this model can handle compounds inviting property or hybrid interpretations. Such compounds constitute an important part of language and should not be ignored (cf. [29]).

3 CoBlLex: Architecture and Operation

Our review of conceptual combination programs shows that all of them lack certain key properties. The programs with adequate combination techniques lack adequate knowledge bases. The programs that have an adequate knowledge base lack adequate techniques. In the present program, CoBlLex, we try to use adequate techniques with an adequate knowledge base. In the following subsections, we sketch the knowledge bases used by CoBlLex and its main combination techniques.

3.1 Knowledge Bases

CoBlLex makes use of four distinct sources of knowledge in its understanding of compounds. First, it uses the WordNet [21] ontology (v. 2.0) - a lexical reference system widely used for language and cognitive engineering applications. Specifically, it uses (i) the synset information of the WordNet corpus, (ii) a gloss for each labelled synset, (iii) hypernym relations between the first and the second word in the compounds, (iv) morphosyntactic, derivative relations between two synsets, (v) generic sentence frames for one or all verbs in a synset illustrating the types of simple sentences in which the verbs in a synset can be used, and (iv) bidirectional attribute relations between noun and adjective synset pairs. CoBlLex also uses eXtended WordNet [11] (v. 2.0) to enhance the information obtained from WordNet. In eXtended WordNet, glosses are syntactically parsed, transformed into logic forms and content words are semantically disambiguated with a view to exploiting the rich information contained in the definitional glosses.

CoBlLex's second source of knowledge is based on CoreLex [2] (v. 1.0). CoreLex defines a set of underspecified semantic types that are further divided to classes exhibiting systematic polysemy. These underspecified semantic types are represented as qualia structures along the lines of Generative Lexicon theory [24]. CoreLex produces a categorization of 39,937 WordNet (v. 1.5) noun instances based on computing a similarity score for the attributes of the lexical items. As such, CoreLex provides a more coarse-grained clustering of word senses, making WordNet data easier to use. The system also offers a unified approach to systematic polysemy of words that is known to play an important part in compound interpretations (cf., [5]). In particular, CoBlLex uses currently about 10% of the CoreLex database of nouns along with 65 nouns (e.g., *stick*, *mouse*, *heart*) that have previously been used in a variety of experimental tests of compounding (see [28]) and are present in WordNet but not in CoreLex. It was necessary to also add 10 new polysemous classes to accommodate these additional nouns.

CoBlLex's third source of knowledge is a collection of familiar or lexicalized compounds not found in WordNet. These compounds were largely collected from empirical studies of human compound use and were verified as being familiar in rating studies (e.g., [28]) and dictionary searches. We included this knowledge because we wanted to allow CoBlLex to distinguish between familiar and novel compounds. Part of our future research program is to leverage the interpretation of novel compounds using familiar compounds (e.g., where someone understands *palmtop computer* with reference to *laptop computer*).

Finally, CoBlLex uses a small number of handcrafted world knowledge rules far and beyond those existing in or derivable from the system sources of knowledge. They are motivated by pragmatic considerations. For example, an *artifact* has a creator. Such rules play a key role in either filtering out implausible interpretations or in allowing plausible interpretations.

3.2 Operation

CoBlLex is fully implemented in SWI-Prolog (v. 5.2). It has a simple interface that asks the user to enter the modifier and head nouns of the compound. This input is parsed and then checked against its knowledge base. If both nouns are known to CoBlLex, then it proceeds to check whether it is a familiar or novel compound. If the compound is familiar, its stored interpretations are retrieved instantly from the database (though there may be only one). No novel interpretations are generated for familiar compounds. If the compound is novel, one or more interpretations have to be constructed from scratch.

The CoBlLex program has a number of core combination techniques that it applies to the knowledge base described above inspired by the previous programs

[14, 5, 17] and designed to capture the variety of types of interpretations known to be produced by people. When people interpret novel, noun-noun compounds they tend to produce three main types of interpretation: relational (e.g., *voice vote*), property (e.g., *octopus table*) and hybrid (e.g. *pear onion*) [27].

To produce relational interpretations CoBlLex uses the relational technique. This technique exploits the hypernym nodes of nouns, attribute relations between noun and adjectives synset pairs, sentence frames and a table of relations. It then creates interpretations following known patterns encoded in the knowledge base. The core procedure is described in Algorithm 1. Take for instance the phrase *book bicycle*. The compound has as modifier a noun belonging to the `artifact-communication` CoreLex class and as head a noun belonging to the `artifact` class. Next, the program will examine the hypernyms of the two nouns and will consult the reasoning engine to check the constraints. From previous similar instances of nouns, the system "knows" that *bicycle* is a subordinate of the concept *transport*, a *book* is an *artifact* and some kind-of *object*, a tangible and visible entity, and none of the world knowledge principles are violated. Then a sentence frame that can accommodate the two nouns is found, i.e.,"something X *transports* something Y", and finally, the program fills the sentence slots with the two constituent concepts accordingly. Notably, the head and the modifier have different roles in a noun-noun pair. Therefore, the order matters. Thus, a *bicycle book* will be interpreted according to a different pattern ("book about bicycles").

To produce property interpretations CoBlLex uses the property technique, which is essentially a mapping of diagnostic features from one concept to the other - typically from the modifier (MN) to the head noun (HN) - as described in Algorithm 2. For example, when the word *demon*, which is member of the `agent-human` CoreLex class, is combined with the word *woman*, a member of the `social_group-human` class, the modifier's glosses are parsed, attributes are extracted and transferred to the head. Possible results are a "woman who is wicked", a "woman who is wicked, skillful and cruel", etc.

Hybrid interpretations are variants of the property interpretations. They are mainly based on a powerset-theoretic model as described in Algorithm 3.

4 CoBlLex: Evaluation

In this section we evaluate the system's performance focusing in particular on how creativity is reflected in the interpretations produced by the methods outlined above.

4.1 Sensibility

To test CoBlLex we generated a set of 100 novel noun-noun pairs that spanned 15 different polysemous classes. We have currently encoded in the system 40 nouns belonging to one of the 15 noun classes and some of their possible combinations. These nouns had been used in a number of psychological studies (e.g., [28]), so

Algorithm 1 Relational Interpretation

```
if MN ∈ CL class artifact-communication and HN ∈ CL class artifact (e.g.,
book bicycle) then
Require:   ∀ d ∈ D → ⊥ | D := world knowledge domain
  if set of conceptual blending rules for relational compounds ≠ ∅ then
    get MN and HN's WN synset chains
    put synsets in lists L_1 and L_2
    look for other diagnostic synsets (DS) in the WN hypernym chain
    if found {E.g., MN isa physical_object ∧ HN isa transport} and MN and/or
    HN derive(s) from WN verb then
      while condition holds do
        retrieve appropriate WN sentence frame
        define roles for MN and HN
        get interpretation I_i
      end while
    end if
    examine table of relations holding between concepts
    if related pattern found then
      while condition holds do
        define roles for MN and HN
        get interpretation I_j
      end while
    end if
    get all interpretations (n = I_1 ... I_k) (e.g., "a bicycle to transport books")
  else
    print WN definitions and a default sentence frame (MN for/about HN)
  end if
  prompt for new entries/exit
else
  examine where MN and HN belong to
end if
```

we had a good idea of how they are typically interpreted by people. Of the 100 presented compounds, 7 were deemed by the system to be uninterpreted because they violated semantic constraints (e.g., *canto car*). In itself, this is an important finding because we know that people will often reach similar conclusions about novel compounds. The remaining 93 compounds were successfully interpreted with CoBlLex producing from 1 up to 1,785 interpretations for each compound. The interpretations were evaluated by two independent raters. Of the total set of approximately 7,000 interpretations produced, 80% on average were deemed to be possible and sensible. To the best of our knowledge, this performance is considerably better than that found in similar previous systems (e.g., [3]).

4.2 Creativity

The essence of CoBlLex's operation is that it knows how to handle the combination of representative noun members of each class and, so, can handle unseen

Algorithm 2 Property Interpretation

if MN $\in$ CL class `agent-human` **and** HN $\in$ CL class `social_group-human` (e.g., *demon woman*) **then**
Require: $\forall\ d \in D \rightarrow \bot \mid D :=$ world knowledge domain
parse MN's and HN's XWN glosses
if diagnostic features (e.g., adjectives) found **then**
put diagnostic features in lists L_1 and L_2
remove duplicate features
generate all interpretations ($n = \sum_{k=1}^{L_1} \binom{L_1}{k}$) (e.g.," a woman who is wicked, skillful, cruel")
else
print WN definitions and a default sentence frame (HN *is like* MN)
end if
prompt for new entries/exit
else
examine where MN and HN belong to
end if

compounds by analogy. The hypothesis is that, if an interpretation is known to be possible between two nouns belonging to certain classes and no world knowledge rules or constraints are violated, then other similar members of these classes may be interpreted in the same fashion. Analogy is an example of combinatorial creativity [1]. It can be seen as a flexible problem solving method [13]. It takes into account attributes matching and case relations between the items in the knowledge base and the target items. Given that in our system deterministic rules apply, we might term this creativity type as *constrained, rule-based combinatorial creativity.*

Although the idea of a machine exhibiting creativity by following a set of rules and constraint-guided inductive methods seems contradictory, this is not necessarily so, as Schank & Owens [26] have argued. Machines have sets of processes and steps that can account for types of creative thinking observed in people. As one example, let us take one of the theories of creativity suggesting that more and less creative people differ on the psychological level, and see whether we can draw any meaningful analogies between such a theoretical model and CoBlLex. Mednick [20] proposed that creative individuals possess the ability to make many and remote associations when a stimulus is presented. On the other hand, less creative people have a relatively "steep hierarchy" - the mental representation of a stimulus is bound to a few other mental representations. If we accept Mednick's hypothesis, then CoBlLex performs indeed like a creative agent. It takes into account polysemy of nouns, exploits the hypernym chains of all concepts (synsets) and word glosses of both nouns, and all the applicable rules before it "decides" what interpretation(s) might be more appropriate for a novel, composite concept. The exploration of different conceptual combinations results often in multiple interpretations. This is exactly what people do when they are asked to interpret novel composite concepts [4].

Algorithm 3 Hybrid Interpretation

if MN $\in$ CL class `food-plant` **and** HN $\in$ CL class `food-plant` (e.g., *pear onion*) **then**

Require: $\forall\ d \in D \rightarrow \bot \mid D :=$ world knowledge domain

parse MN's and HN's XWN glosses

if diagnostic features (e.g., adjectives) found **then**

put diagnostic features in lists L_1 and L_2

remove duplicate features

generate all interpretations ($n = \sum_{k=1}^{L_1} \binom{L_1}{k} * \sum_{l=1}^{L_2} \binom{L_2}{l} - \sum_{m=0}^{L_1 \cap L_2} \binom{L_1 \cap L_2}{m}$)

(e.g.,"a plant that is sweet, juicy, gritty-textured, aromatic, bulbous")

else

print WN definitions and a default sentence frame (HN *and* MN)

end if

prompt for new entries/exit

else

examine where MN and HN belong to

end if

To illustrate all these consider the compound *chocolate car*. A *chocolate car* can be a "car made of chocolate", because WordNet tell us that *chocolate* is a *material* that can be *solid* at room temperature and CoBlLex "knows" that artifacts can be *made of* solid materials. A *chocolate car* can also be a "car with a brown color" because *chocolate* is a kind of *color*, *car* is a *material* artifact and material artifacts can be colored. Following similar chains of reasoning, a *chocolate car* might be a "car that carries chocolate". As part of a psychology pilot study 32 people were asked what a *chocolate car* might be. Almost all of them provided the first two definitions but only 7 mentioned that it might be a "car to transfer chocolate". When these 7 people were asked how they came up with this definition all of them replied that they haven't actually seen a car specifically used to transfer chocolate but they have seen *ice-cream cars* and cars transporting food supplies and, thus, this seemed to be a sensible meaning. Retrieving information about known concept instances and exploring ideas about other similar concepts is a creative action (see [10]) and this is what was observed when people explained their reasoning.

Interestingly, with a different but conceptually similar pairs of nouns people tended to have greater difficulty to find a meaningful link between the two concepts. A *carrot bicycle* was judged by approximately half of the people ($n = 15$) as nonsensical. In terms of the Geneplore model [10] this is not surprising. Geneplore's central proposal is that a creative activity (sense generation is such an activity) can be described with an initial *generative* synthesis of components. Common types of generative processes include retrieval of existing structures from memory, analogical transfer of information from one domain to another, etc. Generative processes may result in so-called *preinventive* structures of varying ambiguity and various exploratory methods (e.g., metaphorical implications and evaluation of structures from different perspectives) need to come into play to explore the structure and make it meaningful. These methods may succeed or fail depending on individual differences and the type of the stimulus. CoBlLex can generate interpretations for “ambiguous and less meaningful” compounds (e.g., “bicycle with orange color”, “bicycle that carries carrots”) if and only if it has sufficiently rich information and combination rules; i.e., if and only if it has the resources to simulate human’s generative and exploratory processes. Then it can perform equally well and sometimes better than an average human (although the system itself will not rank the output by novelty or usefulness). In the authors’ mind, any endeavor to satisfactorily simulate on a machine a living man’s history, experiences and motivational factors, is unrealistic. It is not impossible, though, to enrich representation of concepts and associations between them, and have creative meanings generated for a subset of the ontology domain that humans would miss to find due to the enormous number and complexity of inferences need to be drawn within a very large search space.

4.3 Computational Intensity

The simulation of sense generation for noun-noun pairs has been performed on an Intel Pentium 4 PC (2.8GHz processor with hyper-threading, main memory size of 2Gb, running FreeBSD 5.2 OS). The execution time for finding all possible interpretations ranged between 0.03 and 3 seconds. Prolog’s backward-chaining inference engine operates on the data bases of facts and rules to satisfy the constraints. A number of optimization techniques employed, such as atom garbage collection, recursion optimization, clause indexing, compilation of files to memory and quick loading of object files, significantly speed up the program performance. For novel compounds that had meanings constructed successfully, the number of computational inferences drawn, including the subsidiary (but computationally intensive) natural language processing tasks, ranged between approximately 30,000 and 95,000.

5 Conclusion

Though CoBlLex’s performance is quite impressive it is clearly not perfect. An analysis of the interpretations it produced suggests that its success is heavily dependent on the number of distinct senses captured by the CoreLex semantic class the nouns belong to, and the size and homogeneity of the classes. In general, small classes with fewer and not very polysemous nouns are well defined. When two nouns from such small classes are combined a small number of highly accurate and acceptable interpretations are produced. In contrast, when noun classes are larger the quality of the interpretations produced tends to degrade. In these cases, to be more successful, extra information needs to be induced to sub-cluster the classes members into more finely-distinguished groups. Indeed, some CoreLex classes (e.g., artifacts) are so large and non-homogenous they are almost useless (see also [23]).

The second major determinant of CoBlLex's success is the quality of the WordNet gloss. For example, when producing property or hybrid interpretations, CoBlLex retrieves adjectives associated with the noun to try to find relevant features of the concept. Most of the time this is a useful step but sometimes it finds adjectives that only play an auxiliary role and tell us little about the noun. If occurrence of these non-relevant adjectives could be reduced in some generic fashion then CoBlLex would perform much better in its interpretations. It is interesting to note that sometimes these "noisy" data can give rise to really creative interpretations. Taking again the *demon woman* example, we find among the set of interpretations a "woman who is Christian" and "a woman who is Jewish" just because *Christian* and *Jewish* are annotated as adjectives and they are considered to be diagnostic and informative, by default. It is not hard to imagine social situations were *demon woman* might be used as an insult against a religious group of people. Thus, such interpretations can be accepted. However, the problem with the "noisy" data remains. With the existing generic structure of data that WordNet offers and the syntactical annotation of eXtended WordNet alone, appears to be practically impossible to distinguish what is a diagnostic feature and what is not. Heuristics or extra rules need to be added to constrain the selection of features. Obviously, this is a demanding and not-trivial cognitive engineering task that is open for future research.

Overall, however, we believe that CoBlLex's success is encouraging. It shows that a program with automated word annotation and a small number of handcrafted rules can with limited guidance produce many plausible interpretations and be creative, if we conceive creativity as the production of something *original* and *appropriate* in general [18]. The system development will continue with modelling a larger number of compounds of nouns of different types. We are aware of the difficulties in coping with both semantic flexibility and specificity (see [22]). The core problem in comprehending novel compounds is the interplay between concept partiality, lexical polysemy, ontological assumptions, distributional information, and interpretation patterns stored in memory. We believe that CoBlLex provides a potentially productive approach to understanding the complexities of such interactions; so much so we like to call them "CoBlLexities".

References

1. Boden, M.A.: Computer models of creativity. In Sternberd, R.J., ed.: Handbook of creativity. Cambridge University Press, Cambridge, UK (1999) 351–372
2. Buitellar, P.: CoreLex: Systematic Polysemy and Underspecification. PhD thesis, Department of Computer Science, Brandeis University, Boston (1998)
3. Cater, A., McLoughlin, D.: Noun interpretation using taxonomic links: An experiment in progress. In Monaghan, A., ed.: Proceedings of the Fifth International Conference on Cognitive Science and Natural Language Processing, Dublin City University (1996)
4. Costello, F.J., Keane, M.T.: Polysemy in conceptual combination: Testing the constraint theory of combination. In Shafto, M.G., Langley, P., eds.: Proceedings

of the Nineteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ, Lawrence Erlbaum Associates (1997) 137–142
5. Costello, F.J., Keane, M.T.: Efficient creativity: Constraint-guided conceptual combination. Cognitive Science **24** (2000) 299–349
6. Costello, F.J., Keane, M.T.: Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. Journal of Experimental Psychology: Learning, Memory, and Cognition **27** (2001) 255–271
7. Coulson, S.: Semantic Leaps: Frame Shifting and Conceptual Blending in Meaning Construction. Cambridge University Press, New York and Cambridge (2001)
8. Downing, P.: On the creation and use of english compound nouns. Language **53** (1977) 810–842
9. Fabre, C., Sébillot, P.: Calculability of the semantics of english nominal compounds: Combining general linguistic rules and corpus-based semantic information. Technical Report RR-2742, INRIA (1995)
10. Finke, R.A., Ward, T.B., Smith, S.M.: Creative Cognition: Theory, Research, and Applications. MIT Press, Cambridge, MA and London (1992)
11. Harabagiu, S.M., Miller, G., Moldovan, D.: WordNet 2 - a morphologically and semantically enhanced resource. In Palmer, M., ed.: Proceedings of the Special Interest Group of the Association for Computational Linguistics (1999) 1–8
12. Hayes, J., Veale, T., Seco, N.: Interpreting noun-noun compounds: Exploiting lexical resources to create a truly large-scale model. In Veale, T., ed.: Proceedings of 1st Workshop on Language Resources for Linguistic Creativity, New University of Lisbon (2004) 25–31
13. Keane, M.T.: Analogical Problem Solving. Ellis Horwood Ltd., Chichester, UK (1988)
14. Leonard, R.: The Interpretation of English Noun Sequences on the Computer. North-Holland Linguistic Studies, Amsterdam (1984)
15. Levi, J.: The Syntax and Semantics of Complex Nominals. Academic Press, New York (1978)
16. Lynott, D., Keane, M.T.: PUNC: A model of conceptual combination. In Cunningham, P., Fernando, T., Vogel, C., eds.: Proceedings of the 14th Irish Conference on Artificial Intelligence and Cognitive Science, Trinity College Dublin (2003) 122–127
17. Lynott, D., Tagalakis, G., Keane, M.T.: Conceptual Combination with PUNC. Artificial Intelligence Review (in press)
18. Martindale, C.: Cognition and Consciousness. Dorsey Press, Homewood, IL (1981)
19. McLoughlin, D.: English compound noun interpretation using a lexical database. Master's thesis, Department of Computer Science, University College Dublin (1999)
20. Mednick, S.A.: The associative basis of the creative process. Psychological Review **69** (1962) 220–232
21. Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: An on-line lexical database. International Journal of Lexicography **3** (1990) 235–244
22. Myers, T., Franks, B., Braisby, N.: Partiality and coherence in concept combination. In Dunbar, G., Franks, B., Myers, T., eds.: Edinburgh Working Papers in Cognitive Science: Papers from the Edinburgh Round Table on the Mental Lexicon. Volume 4., Centre for Cognitive Science, University of Edinburgh (1989) 1–19
23. Peters, W., Peters, I., Vossen, P.: Automatic sense clustering in EuroWordNet. In Rubio, A., Gallardo, N., Castro, R., Tejada, A., eds.: Proceedings of the First

International Conference on Language Resources and Evaluation. Volume 1. (1998) 409–416
24. Pustejovsky, J.: The Generative Lexicon. MIT Press, Cambridge, MA (1995)
25. Sampson, S.: English for the Computer: The SUSANNE Corpus and Analytic Scheme. Clarendon Press, Oxford (1995)
26. Schank, R., Owens, C.: The mechanics of creativity. In Kurzweil, R., ed.: The Age of Intelligent Machines. 2 edn. MIT Press, Cambridge, MA (1990) 351–379
27. Shoben, E.J., Gagné, C.L.: Thematic relations and the creation of combined concepts. In Ward, T.B., Smith, S.M., Vaid, J., eds.: Creative Thought: An Investigation of Conceptual Structures and Processes. American Psychological Association, Washington, DC (1997) 31–50
28. Tagalakis, G., Keane, M.T.: Modelling the understanding of noun-noun compounds: The role of familiarity. In Schmalhofer, F., Young, R.M., Katz, G., eds.: Proceedings of the European Cognitive Science Conference 2003, Mahwah, NJ, Lawrence Erlbaum Associates (2003) 319–324
29. Wisniewski, E.J., Love, B.C.: Relations versus properties in conceptual combination. Journal of Memory and Language **38** (1998) 177–202

The Bible is the Christian-Koran: exploring lexical analogy via WordNet

Jer Hayes and Tony Veale and Nuno Seco

Dept. of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland

Abstract. In this paper we examine the phenomena of lexical analogy in terms of analogical compounds. A sub-group of noun-noun compounds appear to be analogical in nature, e.g. the compound "Christian Koran" exists in an analogical relationship with the concept Bible. We outline the properties of these analogical compounds, e.g. the head concept is often a paragon of the category to which the analogue belongs and the modifier is an alignable difference. We also outline a method for discovering these creative concepts from WordNet and how these concepts can involve two distinct types of similarity. Taking these compounds we outline how they can act as seeds for larger analogies in WordNet.

1 Introduction

The process of analogy and its outputs can range in complexity from the verbal analogy tests of the scholastic aptitude test (SAT) to Shakespearean metaphors. We will confine ourselves to a type of compound which appears to be analogical in nature. These compounds raise important questions for concept combination, the process whereby nominal compounds are interpreted. If certain compounds are analogical then perhaps they fall outside of concept combination. Some attempts have been made to account for both literal and figurative compounds [4] but in this paper we adopt an approach where we create analogical compounds from an ontology. The newly generated analogical compounds act as references to existing concepts in the ontology with the references being creative in nature.

A compound is composed of a modifier concept and a head concept [9]. A literal compound is one in which the compound is a hyponym of the head concept. For example a *clothing-store* is a type of store. If we accept that *Christian-Koran* refers to the concept *Bible* then this compound is not literal, rather the head stands for another related category, in this case *sacred-texts*. The head acts as paragon or best example of this category. There is another member of this same category which has the property *Christian*. The modifier therefore marks out a significant property of the analogue and which should be an alignable difference between the source and the target. The major distinction between a literal compound and an analogical compound is that the head in a literal compound refers directly to itself, whereas in an analogical compound the head appears to stand for the general category (hypernym) it belongs to. We suggest that analogies combine literal similarities with alignable differences. In the context of WordNet

(WN) similarities and differences will be determined relative to the word content of a gloss. Literal similarities will be non-trivial overlaps in the gloss (e.g., sharing the property *sacred*) or common hyponyms (e.g., both source and target are *sacred-texts*). Analogical differences will be recognized lexically as a change from one word to another with a lexical affinity, as in from *Muslim* to *Christian*.

We suggest that in principle two types of similarity may be found between a source and target concept in the glosses of WordNet (WN) [8]:

– (1) literal similarities
– (2) alignable differences

In the first case both concepts reference the same concept. This shared feature can act as the basis for the analogy, e.g. the compound *Indian-alder* refers to *kino* as both *alder* and *kino* are related in terms of *tanning*. This can be seen from the glosses for each, *kino* = "East Indian tree yielding a resin or extract often used medicinally and in e.g. tanning" and *alder* = "north temperate shrubs or trees...bark is used in tanning...". The second type of similarity exists between concepts in a gloss which are semantically related. This similarity is known as an alignable difference and it is a difference between corresponding parts of two similar situations or entities [2]. For example, two types of *sacred-texts* may both refer to religion in general but more specifically refer to two distinct religions. The analogical compound *Christian-Koran* refers to *Bible*, with *Koran* listing *Islam* and the *Bible* listing *Christian-religion* respectively in their glosses. Both *Islam* and *Christian-religion* are sibling concepts (they are types of *religion* or *faith*) and are alignable differences. We suggest that the second type of similarity is one that can display the systematicity that analogy is noted for.

2 Creating Analogical compounds

Before we begin outlining our approach to the creation of analogical compounds we suggest that the most unambiguously relevant words in a gloss are the proper nouns. These usually denote nationalities, belief systems, personages and historic/cultural artifacts. We concentrate on these proper nouns and related proper adjectives, recognizing nonetheless that they provide an incomplete picture of the analogical potential of glosses but act as a sound starting point.

Analogical compounds can be created by taking a category which has several children and finding possible analogues between the children where they list possible alignable differences. (These children are also literally similar by virtue of having a common hypernym). For example *sacred-text* has some of the following children {*Veda, Bible*, ...}. An examination of the gloss of each child points to properties associated with each child concept, e.g. *Bible* lists *Christian*. These properties can be used as modifiers to create a new compound with the head being another sibling concept which does not have this property. So *Christian-Veda* could be created and this would refer to the concept from which the modifier was derived, *Bible*. In this paper the properties we will associate with concepts will be proper adjectives. Proper adjectives are used as they are derived from proper nouns and so should be unambiguously relevant words and already act in modifier roles. Proper adjectives can be considered candidate alignable differences.

Taking all concepts in WN, we search for concepts which have more than one child. If a concept A, has children, {X,Y...} and where a child, X, lists a candidate alignable difference (i.e. a proper adjective), P, in its gloss. Then this proper adjective, P, is used as a modifier and forms a new compound with the siblings of A, e.g. P-Y. An analogical compound is thus composed of a modifier which is a property of the target (and an alignable difference) and the head which is the source concept.

2.1 Systematicity

The analogical compounds we have discussed identify how analogy in general may be found in WN. The referencing of the concept *Bible* as the *Christian-Koran* can be seen placing the concepts *Koran* and *Bible* into correspondence. The mappings between both concepts will reflect the two types of similarity we have outlined, literal similarity and alignable differences. A single point correspondence or mapping between *Koran* and *Bible* can give rise to other systematic correspondences or mappings. The gloss of *Koran* lists *Islam* while the gloss of *Bible* lists *Christian-religions* and these new concepts can be used to find other mappings. For example, in any analogical compound one can look for a correspondence between the proper nouns that occur in the gloss of the source and the target and then look for further correspondences in the glosses of these proper nouns. In the case of *Christian-Koran* one would find the following:

Koran : Bible { *Islam : Christian-religions* }, Islam : Christian-religions { *Mohammed : Jesus*

The above lists that *Islam:Christian-religions* was found from a comparison of *Koran* and *Bible* and that in turn a comparison between *Islam* and *Christian-religions* gave rise to a mapping between *Mohammed* and *Jesus*. Also, if we analyse the compound *Christian-religions* we may wish to compare *Islam* to *Christian* if we do we find some of the following mappings:

Islam : Christian { *mosque : church; Koran : Bible* }.

An analogical compound can be used to act as starting point for discovering large systematic analogies that lie undiscovered in WordNet. For example, taking Christian-Koran we could establish a network of mappings which occur between the domains of *Islam* and *Christianity*. This also suggests that of the two types of similarity, a mapping based on alignable differences will be more important in terms of analogy. A mapping based on shared features will be more shallow and perhaps reflect what Gentner has called mere-appearance [1]. However, we should note that just as in any structure-mapping based mechanism (e.g. [11])

we would need mechanisms for ranking mappings and deciding which mappings should be grouped together.

3 External Validation

According to Seco et al [10] one way of validating a newly created compound, when this compound has been generated from an ontology, is external validation. External validation is based on testing the existence of the compound outside of the ontology, for example a compound could be generated which does not exist in the ontology but which exists in a number of web-based documents. The web has been used as a corpus for a number of traditional NLP tasks, e.g. example-based machine translation [12], statistical-based translation [6] and likewise we use the web as a corpus for validating new compounds. This web strategy is motivated by the finding that the number of documents containing a novel compound reliably predicts the human plausibility scores for the compound [5]. Essentially, external validation uses web-based documents to ascertain if the new compound already exists. Given a new compound we submit it to the AltaVista search engine and record how many sites this new compound appears in. This submission looks for the exact phrase of the compound within documents.

4 Results

Taking the mechanism of analogical compound creation outlined in Section 2. We searched through all of WN 2.0 and applied the mechanism to any noun concept which had more than two children. In total 211,085 unique analogical compounds were created. These compounds were then analysed in terms of the two types of similarity between the source and target glosses, (1) literal similarity and (2) alignable differences, and the effect these types of similarity have on external validation. In Figure 1 and Figure 2 external validation is measured against four different thresholds from 40 to 10. A compound which has been externally validated at threshold 40 has appeared in 40+ web documents.

Of the 211,085 analogical compounds created only 19,058 had at least one dimension of similarity (1). However, as Figure 1 suggests the higher the rate of similarity the more probable a compound was to be externally validated. With regards to similarity (2) 102,882 analogical compounds have at least one semantic correspondence between the related source and target. However, the growth in external validation is much more erratic.

One problem in judging the semantic correspondences between a source and a target is that in terms of the analogy in most cases not all the mappings should be counted. Rather those mappings which are systematic should only be counted. This is not the case with Figure 2. However, from Figure 2 it can be seen that prior to point 18 external validation increases. Perhaps after this point several one-to-many mappings occur.

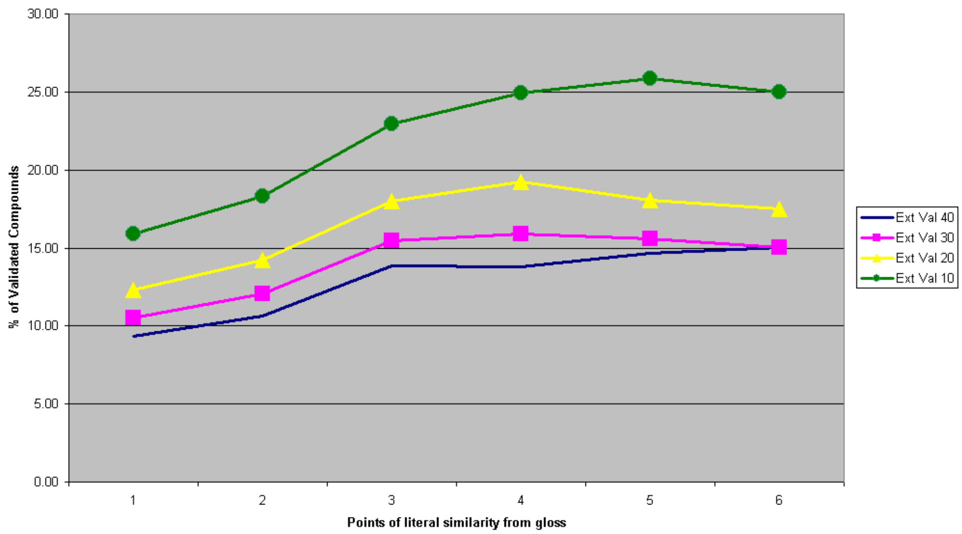

Fig. 1. External Validation in relation to similarity(1)

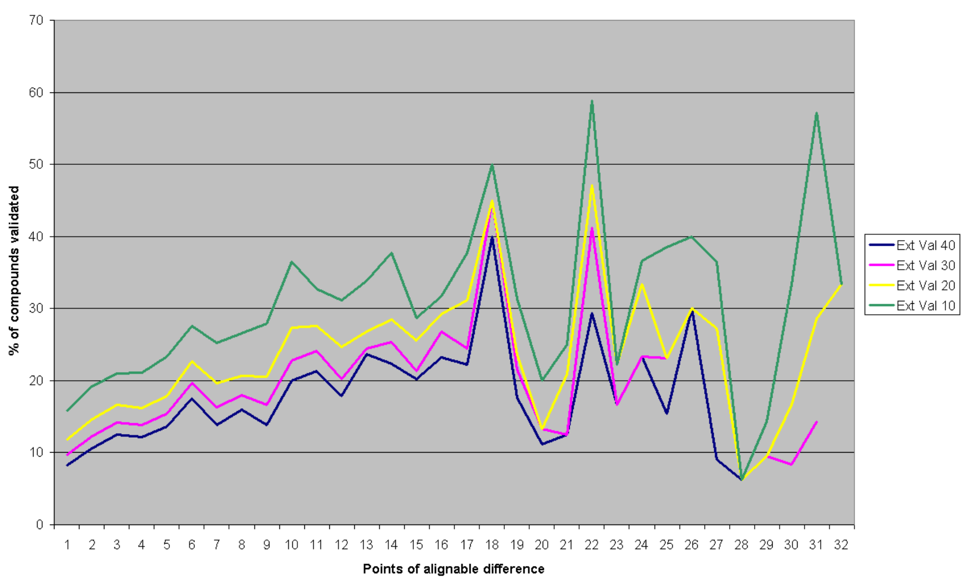

Fig. 2. External Validation in relation to similarity(2)

5 Conclusions and Future Work

We have outlined a type of compound which we have stated is analogical in nature. Taking an ontology such as WN we have outlined a method for creating new analogical compounds from WN (Section 2). We also set out two types of similarity which form the basis for analogy. Taking the external validation technique we outlined how these two distinct types of similarity impact on external validation.

The analysis of simple analogical compounds set out in this paper was carried using only proper adjectives. We intend to examine simple analogies using more than just proper adjectives and intend to explore more complex analogies. One possibility is to use all concepts listed in the glosses of compounds. These could be found by using extended WordNet [7]. The second type of similarity, semantic correspondences, suggests the possibility of using analogical compounds as seeds for larger analogies. These larger analogies would be based on systematic groupings of semantic correspondences.

References

1. Gentner, D. (1987) Mechanisms of Analogical Learning. University of Illinois. Computer Science Department. Number UIUCDCS-R-87-1381.
2. Gentner, D. and Markman, A.B. (1994) Structural alignment in comparison: No difference without similarity. Psychological Science, 5 (3), 153-158.
3. Hanks, P. (2004). WordNet: What is to be done? In the proceedings of GWC'2004, the 2nd Global WordNet conference, Masaryk University, Brno.
4. Hayes, J. (2003). A structural alignment model of noun-noun compound interpretation. Msc thesis, Dublin City University, Ireland.
5. Keller, F. and Lapata, M. (2003). Using the web to obtain frequencies for unseen bigrams. Computational Linguistics.
6. Kraaij, W., Nie, J. and Simard, M. (2003) Embedding Web-Based Statistical Translation Models in Cross-Language Information Retrieval. Computational Linguistics, 29(3), 381–419.
7. R. Mihalcea and D. I. Moldovan. (2001) eXtended Wordnet: progress report. In NAACL 2001 - Workshop on WordNet and Other Lexical Resources, Pittsburgh, PA.
8. Miller, G., Beckwith, R., Fellbaum, C., Gross, D. and Miller. K.J. (1990). Introduction to WordNet: an on-line lexical database. International Journal of Lexicography, 3(4), 35 - 244.
9. O'Grady, W., Dobrovolsky, M. and Aronoff, M. (1993) Contemporary linguistics : an introduction. Longman, New York.
10. Seco, N. and Veale, T. and Hayes, J. (2004) Concept creation in Lexical Ontoligies. In the proceedings of LREC'2004, the 4th international conference on Language Resources and Evaluation. Lisbon, Portugal.
11. Veale, T. (1995) Metaphor, Memory and Meaning: Symbolic and Connectionist Issues in Metaphor Interpretation. PhD. Thesis, University of Dublin, Dublin, Ireland.
12. Way, A. and Gough, N. (2003) wEBMT: Developing and Validating an Example-Based MT System using the World Wide Web. Computational Linguistics, 29(3), 421–457.

Enhancing Sound Design with Conceptual Blending of Sound Descriptors

João M. Martins[1], Francisco C. Pereira[1], Eduardo Reck Miranda[2], and Amílcar Cardoso[1]

[1] AILab, Centro de Informática e Sistemas da Universidade de Coimbra (CISUC)
3030 Coimbra, Portugal
{jpmm, camara, amilcar}@dei.uc.pt
http://ailab.dei.uc.pt

[2] Computer Music Research
University of Plymouth
Plymouth, UK
eduardo.miranda@plymouth.ac.uk
http://neuromusic.soc.plymouth.ac.uk/ccmr.html

Abstract. This paper introduces a new method for sound synthesis using concept description of sounds. Sound descriptions are *blended* to form a new description, which inherits properties from the former entities as well as having an emergent structure of its own. Such blends are then synthesised to become a potentially new sound.
This work applies the system Divago, which is based on a general purpose computational model of Conceptual Blending.

1 Introduction

Computer sound synthesis has become very attractive for a wide range of musicians. Computers are highly programmable and personal computers can run software capable of synthesising sounds in real-time using a wide range of different techniques. Musicians often may not wish to use preset timbres but would rather prefer to create their own instruments. There are, however, a number of ways to implement synthesisers on a computer, and the choice of a suitable synthesis technique is crucial for effective results. Synthesis techniques may be classified into four categories: loose modelling, spectrum modelling, source modelling and time-based approaches.

Loose modelling techniques tend to provide synthesis parameters that bear little relation to the acoustic world. They are usually based entirely upon conceptual mathematical formulae. It is often difficult to predict the outcome and to explore the potential of a loose model. Frequency modulation (FM) is a typical example of loose modelling [1]. FM is a powerful technique and relatively easy to implement, but difficult to operate because the relationship between a timbre and its respective synthesis parameters is not intuitive; for example, increasing the value of a parameter may not necessarily increase the manifestation of its associated sound qualities.

Source modelling and spectrum modelling attempt to alleviate this problem by providing less obscure synthesis parameters; both support the incorporation of natural acoustic phenomena. The fundamental difference between source and spectrum modelling techniques is that the former tends to model a sound at its source, whilst the latter tends to model a sound at the basilar membrane of the human ear. The implementation of a source model (e.g., a physical model) is not straightforward. But once the model is implemented, the user is confronted with relatively intuitive parameters to operate it.

Spectrum modelling techniques have their origins in Fourier's Theorem and the additive synthesis technique. Fourier's Theorem states that any periodic waveform can be modelled as a sum of partials at various amplitude envelopes and time-varying frequencies. Additive synthesis is accepted as being perhaps the most powerful and flexible spectrum modelling method [2]. Musical timbres are composed of dozens of time-varying partials, including harmonic, non-harmonic and noise components. It would require dozens of oscillators, noise generators and envelopes to simulate musical timbres using the classic additive technique. The specification and control of the parameter values for these components are difficult and time-consuming.

Finally, time-based techniques approach synthesis from a time domain perspective. The parameters of time-based synthesis tend to describe sound evolution and transformation of time-related features; e.g. in terms of time lapses. Examples of time modelling techniques include granular synthesis [3] and sequential waveform composition [4]. But again, musicians tend to not use such techniques because it is difficult to determine the role of their parameters with respect to sound quality.

It is clear that some techniques are more intuitive to operate than others, but the most intuitive ones may not be the most appropriate for producing particular types of sounds. From the point of view of the user, however, the problem with sound synthesis is not so much with the intuition of the parameters of the various techniques, but with having the right tools to aid the creative design process.

Sound design is certainly a complex kind of intelligent behaviour. In attempting to solve a sound design problem, composers need to explore possible solutions by trying out possibilities and investigating their consequences. When synthesising sounds to be used in a composition, composers generally have their own ideas about the possibilities of organizing these sounds into a musical structure. In attempting to obtain the desired sound, the composer needs to explore a variety of possible solutions, trying out those possibilities within his or her personal aesthetic. It is the need to provide better support for this exploratory creative process that has motivated our research work.

Sound synthesis systems normally provide good graphic facilities to design the instruments; e.g. visual programming tools such as Max/MSP [5] and Reaktor [6]. However, such systems do not give support for the exploration of the potential of such instruments. The user is often confronted with a visual interface for setting the various synthesis parameters manually. In these cases, the sound design process normally involves non-systematic and lengthy trial and error practices. We believe that we can improve this scenario by providing Artificial Intelligence (AI) to sound design systems. One approach for doing so is to provide tools for the exploration of synthesis algorithms using high-level conceptual descriptions of sounds, as opposed to low-level parametric specifications.

Early attempts at high-level AI systems for sound design have been proposed by Rolland and Pachet [7], and also by Miranda [8], in his system called ARTIST (ARTificial Intelligence Sound Tool). ARTIST was intended to offer the ability to operate the system in terms of qualitative sound descriptors (e.g., adjectives in English) and intuitive operations rather than in terms of numerical values. The system featured a symbolic representation scheme devised to represent sounds in terms of their perceptual components and relations between them.

Other less AI-oriented attempts at the design of high-level interfaces for synthesisers include [9] [10] [11].

A thorough discussion on the pros and cons of these systems is beyond the scope of this paper. It suffices to say that the main limitation of the system developed by [7] is that it has been designed primarily as an interface for a commercial MIDI keyboard synthesiser manufactured in the mid of the 1990s. As for ARTIST, the robustness of the the knowledge base and inference engine has not been tested on cases combining different synthesis methods. Also, the system does not provide a straightforward solution for dealing with conflicting sound descriptors. Unfortunately neither of these systems have been further developed by the authors.

The present paper proposes a further development of the approach introduced in ARTIST, by taking on board a new computational model for conceptual blending, called Divago. We believe that the concept of conceptual blending is more flexible for representing and manipulating sound attributes than the frame-like representation used in ARTIST.

2 Overview of Divago

Divago is a system that is able to combine (i.e. *blend*) a pair of concepts into a single concept which has a structure of its own. In other words, a *blend* inherits characteristics from the original concepts, but may contain novel characteristics obtained from the process or from a third source (e.g. a rule base, an ontology, a *frame*). Thus it should have *emergent structure*. Divago is a rather large project therefore we will give a general overview of the aspects that are relevant for this paper, leaving out some of its foundations and specificities; readers can find this information elsewhere [12, 13].

2.1 Knowledge Representation

Divago allows several different kinds of knowledge representation (the sound synthesis jargon will be clarified in section 3):

- **Concept maps** describe factual knowledge about a concept. A concept map is essentially a semantic network in which all arcs are binary (i.e. they connect

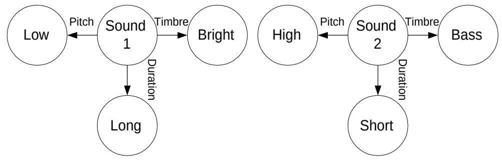

Fig. 1. Concept maps for two sound examples

exactly two different elements[3]). For example, the fact $pitch(sound\_1, low)$ is a characteristic of sound_1 and part of its concept map (see the arc between *sound_1* and *low* in Fig. 1). In Fig. 3 two concept maps of a flute sound and guitar sound are presented as sources for the blend.

– **Rules** describe inferential knowledge about a concept or a domain. Rules are represented in first order logic format[4]. A possible rule could be "If X is a stringed instrument AND it is not a Piano THEN its partials are harmonic.".
– **Frames** describe abstract concepts or procedures. They can be instantiated by the concept maps (when this happens one says that "the frame has been integrated" or "the concept map accomplishes the frame"). They are formally equivalent to rules (their representation is similar). An example of a simple frame could be "wind instrument". If a concept map about a concept c instantiates this frame, then we can say that c is a "wind instrument" (i.e. it would have the generic characteristics expected for the sound of such an instrument, such as a *blow*, $attack - steady - decay$ sequence, etc.). Frames are extremely important in Divago and they can be seen as information moulds which can be used to *shape* new concepts.
– **Integrity constraints** are simple rules (with *false* consequent) that serve to identify inconsistencies (e.g. a sound cannot have a *crescendo* and *diminuendo* at the same time). These constraints, however, do not imply the elimination of the concepts that violate them, rather they are pressures against the generation of these concepts.

2.2 The Architecture

In Fig. 2, we show the architecture of Divago. The **Knowledge Base** comprises a set of concepts (each normally consisting of a concept map) and a *generic domain*, which has generic background knowledge (e.g. an *isa* hierarchy, a set of frames and integrity constraints). For the work presented here, the concepts of the Knowledge Base must be sound descriptions as in Fig. 1. In section 3.1, we will approach this issue in more detail.

The first step for the invention of a new concept (a new sound in the current context) is the selection of the input knowledge, in this case a pair of concepts (e.g. *sound*1 and *sound*2). Currently, this selection is either given by a user or randomly chosen. The **Mapper** then builds a structural alignment between

[3] In order to avoid ambiguity, we call each node of a concept map an *element*.
[4] A rule has the form $C_1 \vee C_2 \vee C_n \Leftarrow P_1 \wedge P_2 \wedge P_m$, for m premises and n conclusions.

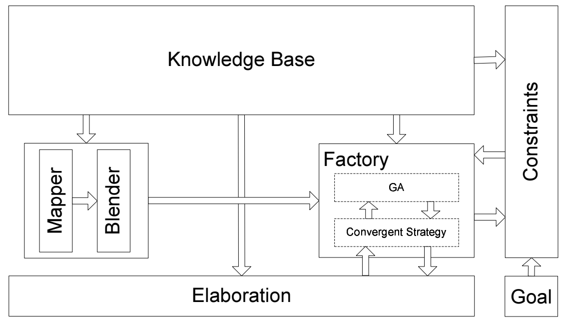

Fig. 2. The architecture of Divago

(the concept maps of) those two concepts. It then passes the resulting mapping to the **Blender**, which then proposes a set of conceptual combinations to be considered, each one corresponding to a *selective projection* from the inputs to the blend. A projection is meant to be the "image" (or the counterpart) in the blend of an element of the input concepts. For example, in the blend of *sound*1 and *sound*2 in Fig. 3, the element *sound*1 gets projected to *sound*2, *bright* gets projected to the blend untouched (*bright*) and *dark* to *bright* (such that, instead of *dark*, *sound*2 will be *bright*). In Fig. 3, we sketch a possible mapping as well as a combination of projections. Notice that not all the elements get projected - a *selective projection* (e.g. *low* in Sound1, *vector*3 in Sound2); an element from the inputs either gets projected to a copy of itself (e.g. *bright* from Sound1), to a copy of its mapping counterpart when it exists (e.g. *dark* from Sound2), or it is not projected at all. Obviously, the number of possible projections is vast for any two input concepts, thus the search space is extremely large. This search space is explored by the Factory module.

The **Factory** is based on a parallel search engine, a *genetic algorithm* (GA), which searches for the blend that best complies with the evaluation given by the Constraints module. Prior to sending each blend to this module, the Factory sends it to the Elaboration module, where it is subject to the application of domain or context-dependent knowledge (in the form of rules and frames found in the generic domain). The GA thus interacts both with the Constraints and Elaboration modules during the search.

The evaluation of a blend given by the **Constraints** module is based on an implementation of the eight Optimality Principles [12], which measure aspects such as *Topology maintenance*, *Frame integration* or *Goal satisfaction*. The **Elaboration** module essentially applies rule-based reasoning (e.g. the application of rules such as the one given in section 2.1). These rules are also part of the knowledge base.

After reaching a satisfactory solution or a specified number of iterations, the Factory stops the GA and returns the best solution achieved, also in the form of a concept map (and with new rules, frames or integrity constraints, in the rare cases in which these structures are also part of the input concepts and of the

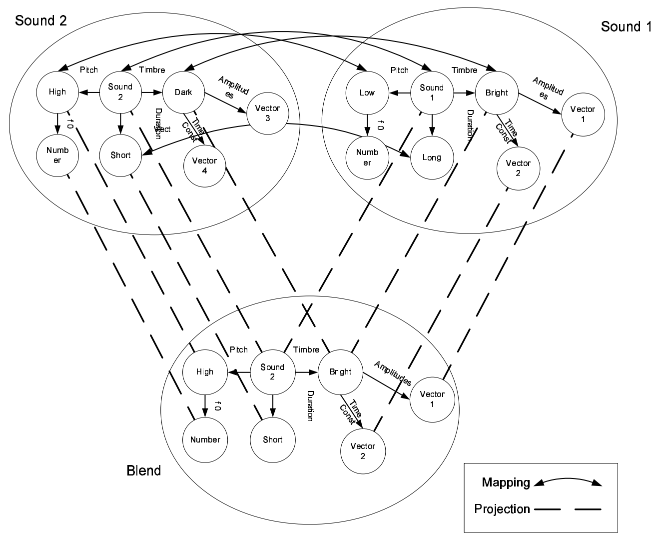

Fig. 3. A blend with its mappings and projections (for the sake of readability, we show only an excerpt).

blend). Thus, the input and output of Divago is expressed in the same syntax and with the same kind of knowledge structures as described in section 2.1.

In some cases, the output of Divago is also the input of an **Interpretation** module, which produces an interpretation of the new concept. In previous versions of this system, we made interpreters for 2D [14] and 3D images [15], as well as textual descriptions of the blend [16]. Of course, these several *modalities* were adapted to specific uses and therefore they are not guaranteed to work in different applications. For the present work, our Interpretation module will correspond to a *Synthesiser*, as will be explained in section 3.3.

3 A Case-Study System

This paper reports an extension of Divago to generate blends of sounds. More specifically, it consists of a knowledge base with sound descriptions, and frames, rules and integrity constraints that are more appropriate to the sound synthesis domain. On the output side, a synthesiser interpreter is being developed as explained in section 3.3.

As a first set of experiments, the concept maps from Fig. 1 were programmed in Divago, as well as integrity constraints for forbidding concurrence of sound states (e.g. a sound cannot be *long* and *short* at the same time). We also gave one frame to the system: $timbre\_transfer$, which expects the blend to have a *timbre* (and its associated vectors) from one input in the context of the other input. When used in the query, this frame values the transfer of *timbres* to a new context. In other words, sounds with the *new timbres* will be preferred by the genetic algorithm. The blend generated (from a set of 30 runs) is described in Fig. 4.

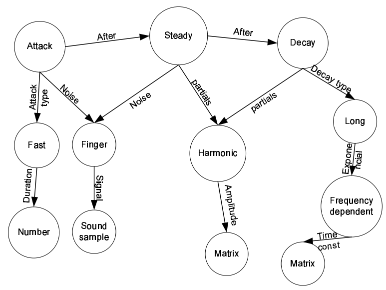

Fig. 4. A blend of sound1 and sound2 generated by Divago

This output thus needs an interpretation, a signal synthesis that results from an unambiguous reading of the blends. In the following subsection we will describe the concepts handled by Divago which will be interpreted by the synthesiser.

3.1 Sound Descriptions

Divago needs descriptions at the concept level, preferably in the form of concept maps. This implies the (always subjective) choice of a language and of abstract level primitives to describe sounds.

Four general characteristics are commonly used to describe sound [17]: *pitch*, *duration*, *timbre* and *loudness*.

All these attributes are subjective, each being dependant on more than one measurable physical characteristic, such as *pressure*, *frequency*, *spectrum*, *envelope*, and *duration*. From these, the least understood are spectrum and envelope. Spectrum is the space where the frequency content of a sound is pictured, and each frequency has a corresponding amplitude. Prominent peaks in spectrum are called partials. Envelope is the time variation of the amplitude (or energy) of sound.

Pitch is the attribute by which sounds can be ordered from low to high, most musical instruments have a defined pitch, except for some percussion instruments. Although pitch depends strongly on the frequency of the fundamental tone, it is also influenced by the intensity of the sound and its high frequency components (spectrum). Loudness is a perceptual measure of the intensity of sound. It depends mainly on sound pressure exerted on the timpani of the ear, but can also be influenced by the spectrum and duration of the sound. The American National Standards Institute (ANSI) defines timbre as "... that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different". This attribute identifies the sound source and is the most tricky attribute to quantify as it depends on many different characteristics. Another problem is that there is no uniform set of concepts to identify and classify timbre. Analogies with visual and tactile expressions are often created to suppress this lack of concepts in the musical domain: sounds can be warm, dark, bright, sweet, metallic, etc. It is certain that timbre shows a strong dependence on spectral components and envelope characteristics. Statistical tests show that the transients of the attack and decay parts are critical for instrument identification [18]. Pollard and Janssen [19] designed a graphic representation method called *Tristimulus*, analogous to that used for mapping colours. In this method the relations between partials are mapped in two dimensions, marking the evolution of the partials in time on a graph, producing a visual representation of timbre. For more detailed description of the perceptual attributes of sound please refer to [20].

To control the previously described attributes we will use additive synthesis, as it is easy to understand and is the basis for more advanced techniques, such as Spectral Modelling. Sounds are synthesised by weighting and adding sinewaves with different frequencies [3] along the spectrum. These sinewaves model the partials of the original sound.

In later experiments we use guitar and flute sounds, dividing them into three distinct parts by inspection of the sound envelope, as shown in Fig.5, resulting in three concepts connected by the concept map: *attack*, *steady state*, and *decay*. We now explain these concepts as Sound1 and Sound2 taking the form of *decay* sounds.

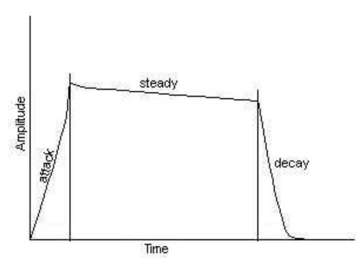

Fig. 5. Envelope model of a musical instrument sound

Attack is the initial part of the sound. It contains noise components from the physical interaction between the player and the musical instrument, as well as resonances from the body of the musical instrument that usually fade away quickly, and the raising partials from the original vibrating source.

Steady state is the signal part corresponding to sound driven by a player. In struck or plucked instruments, sounds do not possess steady parts, as they are not driven by a force or constant blow.

Decay is assumed to be the natural phenomenon of the attenuation of a sound when it is not fed with external energy. Here we keep the same parameters

Partials		$1(f_0)$	2	3	4	5	6
Sound 1	weight	3.5	3.0	2.0	2.0	3.0	1.0
$f_0 = 110$Hz	Time const. (s)	0.5	0.3	0.5	2.0	2.0	2.0
Sound 2	weight	3.0	0.5	0.5	0	0	0
($f_0 = 240$Hz)	Time const. (s)	1/3	1/6	1/9	0	0	0

Table 1. The weights and time constants of the exponentials of the first six partials of sound examples 1 and 2

as in steady state, except for duration, which we substitute by exponential factors for each partial. The decay time is the time between the end of the steady state (or attack in case of instruments without this state) and the time point where the signal decays to $1/e$ of its initial value. In reproducing natural sounds it is crucial to obtain different decays for each partial.The low partials of a string decay more slowly than the higher ones [21].

The experiments carried out with the sounds shown in Fig. 1 lead us to associate two vectors to the concept following the arc *timbre*: amplitudes and time constants of the partials. Although Sound 1 has a low pitch, which is given by its fundamental frequency of 110 Hz, the row vectors from Tab. 1 show that the high amplitudes and long decay times in the high frequencies account for a *bright* timbre. The insignificance of the high frequencies in Sound 2 accounts for its *dark* timbre.

3.2 Other Structures

Frames For the current setting, we use the same general-purpose frames that have been applied in other experiments [13, 22, 15]. The frames $aframe$ and $bframe$ imply the same relational structure as inputs 1 and 2 (resp.), i.e. a blend that integrates these frames will have the same relations as those inputs. The frames *aprojection* and *bprojection* imply the projection of the same elements of inputs 1 and 2 (resp.). In other words, when a blend integrates these frames, the nodes being used come from those inputs. Other frames from previous works could be used and will certainly be subject to experiments, however these four frames are the ones that are context independent and have proven to be fundamental in the construction of blends.

We have also built frames specific to the Sound domain. For the purposes of this paper, the frames for $timbre\_transfer$ as described above and used in the query, for two kinds of sound state sequences ($attack \rightarrow decay$ and $attack \rightarrow steady \rightarrow decay$) and for the two kinds of instruments were coded. Below, we show the coding of the frame $steady\_sound$:

$$frame(steady\_sound : \\ steady\_sound \longleftarrow after(attack, steady) \wedge after(steady, decay)$$

More knowledge could be included, such as the conditions that should be present for a state to be considered: *attack*, *steady* and *decay*. In table 2, we show the frames that are currently available in the knowledge base of Divago.

Frame name	**Conditions**
aframe	The blend contains identical structure from input 1
aprojection	The blend contains the same elements of input 1
bframe	The blend contains identical structure from input 2
bprojection	The blend contains the same elements of input 2
timbre_transfer	The blend results from the transfer of the timbre of one input to the context of the other input
steady_sound	The blend follows the sequence of states *attack* → *steady* → *decay*
attack_decay_sound	The blend follows the sequence of states *attack* → *decay*
wind_instrument	The blend contains the characteristics of a wind instrument
plucked_instrument	The blend contains the characteristics of a plucked instrument

Table 2. Some frames of the generic space

Integrity Constraints As the integrity constraints are essentially domain dependent, we add them as the system progresses in each new domain. For the experiments referred to in this paper, we only used three integrity constraints for avoiding sound state concurrence:

$$false \leftarrow duration(X,Y) \wedge duration(X,Z) \wedge Y \neq Z$$
$$false \leftarrow timbre(X,Y) \wedge timbre(Z,Y) \wedge X \neq Z$$
$$false \leftarrow pitch(X,Y) \wedge pitch(Z,Y) \wedge X \neq Z$$

Goals As for the rest of the knowledge structures in Divago, the language of goals allows the same possibilities any Prolog interpreter can offer, which implies that, when submitting a query to Divago, we can use simple pairs of relations and reference to frames or even entire logic programs. Nevertheless, experience has told us that using frames and simple relations is enough to make Divago give us satisfactory results. For example, for the result shown in Fig. 4, the query contained only $timbre\_transfer$.

3.3 Synthesiser

Spectral and Physical models are synthesis techniques that offer us good perspectives to interpret blended sounds [23], but we leave these for the near future, as we have been using the more simple technique of additive synthesis programmed in Matlab, to illustrate the system. At this point we have used the blend created by Divago, shown in Fig.4 to create a sound that maintains most of the characteristics of Sound2, yet having the timbre deriving from Sound1. The Matlab code and sound examples used to generate this examples can be found at our URL [24].

In other experiments we have blended guitar and flute sounds putting emphasis on the temporal division of sound addressed in Sec. 3.1, using *attack*, *steady* and *decay* in a map of concepts. The extraction of the split points between these regions shown in Fig. 5 is not trivial, and although there are some methods described by Jensen [25] to extract them, we used only visual inspection. The amplitudes of the partials of both the steady state of the flute and the guitar decay, were extracted with an analysis tool developed at the Helsinki University of Technology [26], and have 19 components.

The resulting blend resembles a guitar sound, as it has the same pitch, the same decays and the same spectral distribution. The new feature about it is the existence of a steady state projected by the flute sound, lasting for approximately 2 seconds, with a sampling rate of 22050 Hz.

The blend signal is pictured on Fig. 6. It is important to notice that we are not only creating a new instrument somewhere between the former instruments, but we are also exploring the conceptual description and features of the sound.

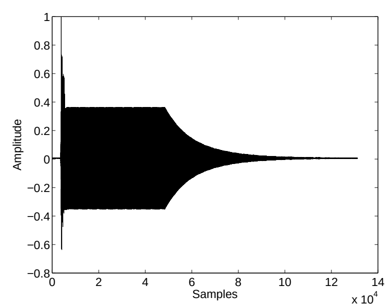

Fig. 6. The result of a guitar-flute blend

4 Conclusions and Further Work

Taking Divago as a basis for a sound synthesis system seems a promising idea as it has demonstrated versatility and creative capabilities in various domains. On one hand, we may test its potential in the domain of sound synthesis, and we may also develop a method for the generation of sounds based on abstract descriptions, designed from perceptual or cognitive perspectives of sound analysis. The only restrictions are that these descriptions should be unambiguous and correspond to Divagos syntax. Currently, we achieve this with a distinct semantics for each element of the concept maps used (e.g. “attack” is interpreted as the attack part of a sound signal). A further goal is the creation of an automatic interpreter to transpose the knowledge from the concept maps to the synthesiser module. Also, the use of Physical and Spectral Modelling remain future goals to explore the concepts associated with the sound production mechanisms.

References

1. Chowning, J., Bristow, D.: FM Theory and Applications: By Musicians for Musicians. Tokyo: Yamaha Music Foundation (1986)
2. Dodge, C., Jerse, T.: Computer Music. New York: Schirmer Books (1985)
3. Miranda, E.R.: Computer Sound Design: Synthesis Techniques and Programming. Oxford (UK): Focal Press (2002)
4. Chandra, A.: The linear change of waveform segments causing non-linear changes of timbral presence. Contemporary Music Review: Timbre Composition in Electroacoustic Music **10** (1994) 157–169
5. Cycling74. http://www.cycling74.com/ (Last visited 6 July 2004)
6. Sasso, L.: Native Instruments Reaktor 3 The ultimate hands-on guide for all Reaktor fans. Bremen (Germany): Wizoo (2001)
7. Rolland, P.Y., Pachet, F.: Representation de connaissances sur la programmation de synthetiseurs. In: Recherches et Applications en Informatique Musicale. Volume 1998. Hermes (Collection Informatique Musicale)
8. Miranda, E.R.: An artificial intelligence approach to sound design. Computer Music Journal **19** (1995) 59–75
9. Ethington, R., Punch, B.: Seawave: A system for musical timbre description. Computer Music Journal **18** (1994) 30–39
10. Garton, B.: The elthar program. Perspectives of New Music **27** (1989) 6–41
11. Schmidt, B.L.: Natural language interfaces and their application to music systems. In: Proc. of the 5th Audio Eng. Soc. International Conference. (1987) 198–206
12. Pereira, F.C., Cardoso, A.: Optimality principles for conceptual blending: A first computational approach. AISB Journal **1** (2003)
13. Pereira, F.C., Cardoso, A.: The horse-bird creature generation experiment. AISB Journal **1** (2003)
14. Pereira, F.C., Cardoso, A.: The boat-house visual blending experience. In: Proc. to the 2nd Workshop on Creative Systems, ECAI'02 (2002)
15. Ribeiro, P., Pereira, F.C., Marques, B., Leitao, B., Cardoso, A.: A model for creativity in creature generation. In: Proc. of the 4^{th} Conference on Games Development (GAME ON'03), EuroSIS / University of Wolverhampton (2003)
16. Pereira, F.C., Gervás, P.: Natural language generation from concept blends. In: AISB'03 Symposium on AI and Creativity in Arts and Science, SSAISB (2003)
17. Rossing, T.: The Science of Sound. Addison-Wesley (1990)
18. Berger, K.W.: Some factors in the recognition of timbre. J. Acoust. Soc. of America **36** (1964)
19. Pollard, H., Jansson, E.: A tristimulus method for the specification of musical timbre. Acustica **51** (1982)
20. Cook, P.: Music, Cognition, and Computerized Sound. The MIT Press (1999)
21. Fletcher, N., Rossing, T.: The Physics of musical instruments. Springer-Verlag New York Inc. (1991)
22. Pereira, F.C.: Experiments with free concept generation in divago. In Cardoso, A., Bento, C., Gero, J., eds.: Proc. of the 3rd Workshop on Creative Systems, IJCAI-03 (2003)

23. Smith, J.O.: Viewpoints on the history of digital synthesis. In: Proc. Int. Computer Music Conf. (ICMC-91). (1991) 1–10
24. Martins, J.M. http://eden.dei.uc.pt/~jpmm/CC04/ (Last visited 6 July 2004)
25. Jensen, K.: Envelope model of isolated musical sounds. In: 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99). (1999)
26. Välimäki, V., Tolonen, T.: Development and calibration of a guitar synthesizer. J. Audio Eng. Soc. **46** (1998) 766–778

The Role of Gatekeepers in Creativity

Ricardo Sosa and John S Gero

Key Centre of Design Computing and Cognition,
University of Sydney, NSW 2006, Australia
r.sosa@usyd.edu.au, john@arch.usyd.edu.au

Abstract. This paper investigates some principles of gatekeeping in creativity. It presents results of a computational framework based on Csikszentmihalyi's DIFI framework [6] applied to design. In this multi-agent system individuals are designers of artifacts, the field is composed by adopters and opinion leaders, and the domain consists of a collective repository of selected artifacts. The adopter population is organized in social networks where adopters influence the adoption decision process of others. Opinion leaders emerge as a result of this form of social interaction and become responsible for selecting entries to the collective repository. With these simple elements, a number of interesting phenomena are demonstrated in relation to gatekeeping in creativity. The findings suggest that the emergence of creative figures can be better understood when situational factors are considered, such as the role of the field.

1 Introduction

In this paper creativity is defined as a property socially ascribed to individuals that generate solutions considered as novel and useful by members of their society. Most definitions in the literature acknowledge this relationship between generative and evaluative processes. Historical or H-creativity is indeed different from everyday creativity in that groups of people agree in their novelty and appropriateness [1]. However, mainstream creativity research has been largely preoccupied with the study of individual characteristics of creators implicitly treating evaluation as the appreciation of talent by a passive audience. Conclusive evidence on assumed *creative traits* (i.e. pertaining exclusively and consistently to creative individuals or processes) is yet to be shown [2–4]. Differences in patterns of creative production as shown in Fig. 1 suggest that if no unique personality or process is correlated to creativity, it could be because creativity occurs within a system rather than "within-the-head" [5].

This paper presents a computational framework of design as a social activity based on the DIFI framework, which maps the main components of the creative system: individuals, fields, and domains [6]. In this paper a fundamental field factor is explored, that of gatekeeping or the social process of granting access to a domain shared by a group. Gatekeepers are considered opinion leaders that emerge from bottom-up social organization. They are assumed to represent some aspects of their social groups and to exert influence over their choices. These influential figures become responsible for the selection of design solutions for their inclusion in the domain. By studying the lives of creative figures like Picasso and Freud, Gardner [7] suggests that the structure of the field could affect how creators become prominent. Evidence of field determinants are also seen in interdisciplinary differences between age and creative output as shown in Fig. 2. This type of field characterization demonstrates that factors outside the individual take place in the definition of creativity.

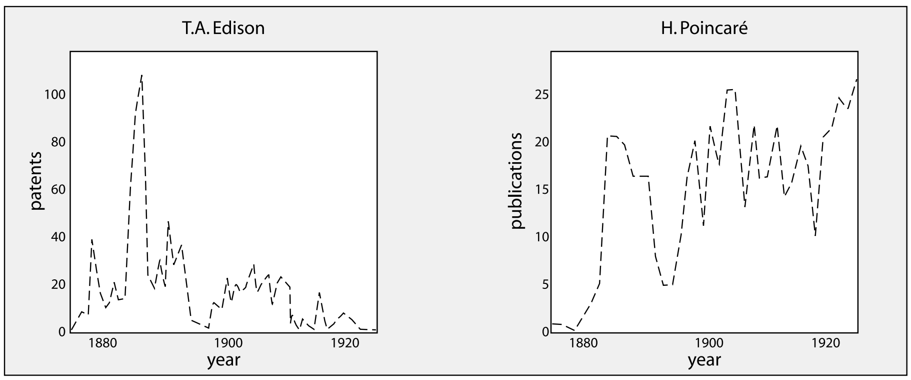

Fig. 1. Comparison of production rates of Thomas Edison and Henri Poincaré illustrates the elusiveness of individualistic characterizations of creativity [4].

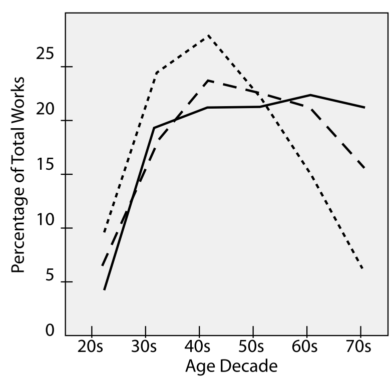

Fig. 2. Total output of creative figures by chronological age indicates interdisciplinary differences. Creative output varies in slope and peak across disciplines [3].

In design, gatekeeping can be studied at the team level or at the social level. Members of a design team may operate both as designers and as adopters or internal clients. At the social level gatekeepers in design include patent examiners, venture capital firms, and competition juries. In both cases gatekeeping results from interaction of opinions and can be seen as the process of evaluating and selecting ideas.

The aim of this research is to extend our understanding of creativity by exploring the effects of situational factors. In particular this paper focuses on the effects that differences in gatekeeping have on the occurrence and recognition of creativity.

2 Framework of Design and Social Influence

2.1 Artifacts

In our framework artifacts are implemented in a two-dimensional line representation in a 3 × 3 grid as shown in Fig. 3a. This is a simple way of representing features of design artifacts with common constraints. Multiple representation and ambiguity are possible because artifacts are perceived and interpreted by adopters according to a set of randomly distributed perception and adoption biases. Figure 3b shows possible perceptions of the artifact in Fig. 3a. The assumption is that people perceive and base their evaluations on different features of design products. At the implementation level, artifacts are coordinate arrays whilst perceived features are closed shapes built as Hamiltonian circuits based on a branch threshold of $\boldsymbol{v} \pm \mathbf{2}$ [9] where $\boldsymbol{v}$ is an integer randomly drawn from a Gaussian distribution with mean and standard deviation as independent variables.

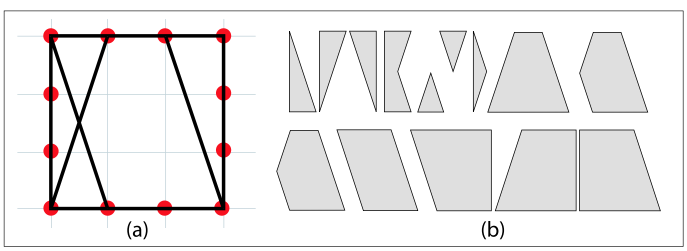

Fig. 3. (a) A simple design artifact representation and (b) possible interpretations of an artifact built as Hamiltonian closed shapes by adopters with individual perception biases.

2.2 Adopters

Adopters are implemented as social agents [10, 11]. The adoption decisions of social adopters are partly determined by individual factors such as preferences and partly by social factors such as influence of opinions. Given their individual sets of perceived features $\boldsymbol{G}$, adopters evaluate artifacts based on a multi-objective adoption function, $\boldsymbol{Afn}$, where six geometric criteria $\boldsymbol{C}$ are evaluated, i.e. alignment $(\boldsymbol{x}, \boldsymbol{y})$, intersection $(\boldsymbol{i})$, rotation $(\boldsymbol{r})$, similar bounds $(\boldsymbol{b})$, and number of sides based on their individual threshold of perception $(\boldsymbol{s})$ as defined in (1).

$$Afn = \left(\frac{\sum[(C_x/G), (C_y/G), (C_i/G), (C_r/G), (C_b/G), (C_s/G)]}{(G^2 - G)} \right). \qquad (1)$$

Where the adoption function $(\boldsymbol{Afn})$ is individually estimated as the sum of geometric relations of perceived features $\boldsymbol{G}$. Therefore, an adopter will assign a higher $\boldsymbol{Afn}$ to artifacts perceived as having more shapes with these geometric relationships. Individual preferences are implemented as biases $(0.0 \leq \boldsymbol{C} \leq 1.0)$ for each geometric criteria $\boldsymbol{C}$, i.e. adopters with preferences for aligned shapes assign an extra weight to that criterion across all available artifacts.

The adoption decision of an agent is given by the highest perceived $\boldsymbol{Afn}$. If an adopter perceives no difference between artifacts, it abstains from adopting. In principle, adopters with different preferences could reach equivalent adoption decisions if their evaluations are based on different perceived features $\boldsymbol{G}$. This can occur if they base their $\boldsymbol{Afn}$ on different perceived attributes of an artifact. Likewise, adopters with similar preferences need not adopt the same artifact if their evaluations are based on different interpretations.

Social interaction complements the adoption decision process. It consists of contact with nearby adopters where the aim is to influence certain aspects of their adoption decisions. Adopters are assigned random positions in different social networks where adopters are represented as nodes and their adjacency by links or social ties [12]. In this paper three social networks or spaces are implemented where each space is defined by the content of interaction. In the first space criteria preferences are exchanged, in the second space artifact features are exchanged and adoption decisions are exchanged in the third social space. Given an extroversion threshold [13, 14] adopters spread these elements of their opinions to other adopters. After a certain period of interaction, a group of adopters may thus emerge that shares similar preferences or similar perceived features of artifacts. Random walks in one and two dimensional spaces are recurrent, i.e. have a probability of 1 of visiting the same point given sufficient time. However, this probability in $\geq$3-dimensional random walks approaches zero [15]. As a result, adopter populations based on three or more social spaces need not reach equilibrium even during long system runs.

2.3 Gatekeepers

Opinion leaders are adopters that gain a position of influence in a social network. In this paper adopters gain the position of opinion leaders when their influence over other adopters is two standard deviations above the population mean. While adopters maintain the role of opinion leaders, their choices influence future adoption decisions of the population by marginally increasing the weight of preferred geometric criteria. More importantly, opinion leaders are responsible of selecting artifacts based on a more comprehensive geometric function that includes symmetry and scale criteria. Namely the entry value of the domain repository function, $\boldsymbol{Dfn}$, where three extra geometric criteria $\boldsymbol{C}$ are evaluated, i.e. symmetries $(\boldsymbol{x', y'})$, and uniform scale $(\boldsymbol{u})$ as defined in (2).

$$Dfn = \left(\frac{\sum[(C_{x'}/G), (C_{y'}/G), (C_u/G)]}{(G^2 - G)} \right). \qquad (2)$$

The domain is implemented as a repository of entries selected periodically by opinion leaders where the rate of selection is an independent variable. In this way they become "gatekeepers" of their social group. As gatekeepers select entries, the threshold of entry is set to the $\boldsymbol{Dfn}$ value of the last entry. Future entries can only be added to the repository if they receive a higher $\boldsymbol{Dfn}$ value by gatekeepers.

2.4 Designers

Designers are assigned equivalent artifacts at initial time of a simulation run. At regular intervals they evaluate and modify their artifacts. This 'design rate' is an independent variable measured by adoption iterations. Namely, the system schedule can be set so that the adopter population evaluates and reaches adoption decisions a variable number of times between every design update. After a design review, adopters re-perceive the modified artifacts and continue their adoption evaluation and social exchange.

Designers contact opinion leaders to model the population's current preferences and perception biases. They evaluate their own artifacts by applying the adopters' $\boldsymbol{Dfn}$. This way a designer estimates what features of its artifact are likely to be perceived and how well these perform. Designers modify the features with lowest ranking, i.e. those with the lowest contribution to $\boldsymbol{Dfn}$ criteria. The designer substitutes the line or lines that produce such feature with a random line. By modeling the decision process of adopters, the designer estimates if such a line increases the $\boldsymbol{Dfn}$ and therefore is likely to increase the future adoption of its artifact. When designers find a line that increases the $\boldsymbol{Dfn}$ of their artifacts, they associate that line with the preferred criterion as a condition-action rule. If no increase occurs after a limit number of trials, the designer selects the leading adopted artifact and substitutes its low ranking line with one random line of that artifact. This imitation process is followed by the acknowledgement to the designer of the source artifact as an increase of peer recognition.

The behavior of a designer can be characterized by the rules or knowledge generated, the entries selected by opinion leaders, the credit given by other designers when imitating, and the size of their adoption bases. These elements model in simple ways phenomena such as popularity [3], peer recognition, expert endorsement, and expertise [16]. In this paper three types of experiments are reported to explore field effects. These experiments explore variations in the strength of social ties, gatekeeping rates and population size. In all cases designers' characteristics are kept constant as a way to assess the role of situational factors.

3 Results

This paper reports the effects of gatekeeping by manipulating three related independent variables. All system runs are initialized with equivalent conditions. Results are mean values of Monte Carlo simulations run for 2500 iterations and 30 cases. The parameter range of each one of the three independent variables is explored with increments of 10, unless otherwise specified. Outliers are defined as data points two standard deviations apart from the mean and are excluded from the dataset.

3.1 Strength of Social Ties ($\boldsymbol{T}$)

Social ties are defined as interaction links between nodes in a social network where nodes represent the location of adopter agents in that particular social space [12]. The strength of social ties, $\boldsymbol{T}$, is determined by the probability that associated nodes interact over a period of time [17]. Strong social ties exist between nodes in a kinship network, whilst weak ties exist in networks where casual encounters occur between strangers or acquaintances. In our framework we implement a basic notion of tie strength as a probability ($0.0 \leq \boldsymbol{T} \leq 1.0$) that the link between a possible pair of adopter agents will remain at the next time step [18]. $\boldsymbol{T} \approx 0.0$ brings higher social mobility, i.e. adopter agents are shuffled more often and get to interact with different adopters over a period of time. In contrast, $\boldsymbol{T} \approx 1.0$ bonds adopters together causing a decrease in social mobility, i.e. adopter agents interact within the same groups for longer periods.

Adopter influence in this framework is measured by the Gini coefficient, a summary statistic of inequality. The Gini coefficient γ measures the distribution of limited resources that are exchanged among members of a population. Influence can be seen as a limited resource in that it is generated by the interaction between an agent pair where one increases its share at the expense of another. When $\gamma \approx 1.0$ influence is concentrated by a few adopters and more stable dominance hierarchies exist. In contrast, when $\gamma \approx 0.0$, influence is more distributed among adopters.

The results of varying $\boldsymbol{T}$ from 0.0 to 1.0 show that as $\boldsymbol{T}$ increases, social mobility decreases causing adopters to interact more often with a stable group of neighbors. As a result, influence is more concentrated ($\gamma \approx 1.0$), i.e. a few adopters exert dominance over others. In contrast, as $\boldsymbol{T}$ decreases, social mobility increases and agents have contact within a varying neighborhood. In such conditions, influence structures of dominance are more distributed ($\gamma \approx 0.0$), i.e. hierarchies become more flat. Figure 4 shows a scatter plot of the power-law relation of tie strength $\boldsymbol{T}$ and Gini coefficient γ. This result suggests the following principle of social ties and influence of opinion: In social groups with strong ties there is lower mobility and the spread of ideas occurs in a hierarchical structure of influence between adopters. As a result, *in groups with strong links influence hierarchies guarantee that a few individuals become dominant in the spread of ideas*. The relation between tie strength $\boldsymbol{T}$ and influence distribution is nonlinear. Social groups with strong ties $\boldsymbol{T} \approx 1.0$ reach a mean Gini coefficient $\gamma = 0.44$. As $\boldsymbol{T}$ decreases marginally, there is a sudden drop of influence hierarchies rapidly going below $\gamma = 0.39$. However, once this threshold is crossed, even a significant decrease in $\boldsymbol{T}$ does not pull $\gamma < 0.38$. This provides a refinement to the principle of social interaction as follows: Whilst strong social ties cause large influence hierarchies, small amounts of social mobility in such societies rapidly reduce disparities. As tie strength decreases further, influence becomes more egalitarian up to a point at which even large changes in social tie strength and mobility do not have a significant impact.

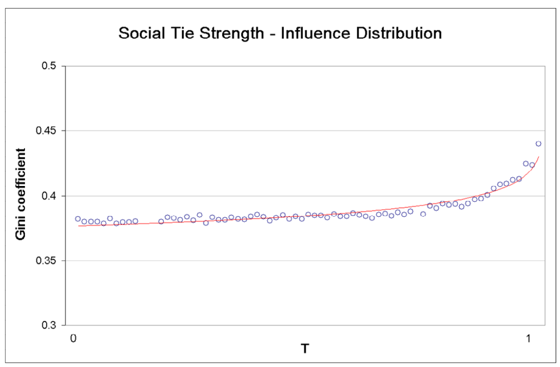

Fig. 4. Relation between tie strength $\boldsymbol{T}$ and influence distribution γ. Strong social ties generate hierarchical structures of influence where a few adopters become very influential. Weak social ties generate more spread influence structures with narrow dominance gaps.

In the domain, a consistent increase in repository size takes place as social ties become weaker as shown in Fig. 5. In societies with influence hierarchies $\boldsymbol{T} \approx 1.0$ the same adopters tend to remain in the role of gatekeepers. Namely, gatekeeping is more stable and controlled by a small unchanging group of influential experts. Therefore, interpretations in which the evaluation of artifacts is based, remain constant over time. Two consequences are that repositories tend to be smaller and entries originate from a lower number of designers.

In contrast, in societies with lower tie strength $\boldsymbol{T}$ and therefore where influence is distributed rather than concentrated there is a higher change rate of gatekeepers. The gatekeeping group is constantly composed of different adopters. Consequently, more diverse evaluations mean more artifacts are submitted to the repository. A principle of tie strength and repositories can be stated as: In fields where social ties are strong and influence is concentrated, an unvarying group of gatekeepers generates smaller domains. Under such conditions, entries to the repository originate from a small number of designers. In fields where social ties are weak and influence is more distributed, there is a high rotation of gatekeepers. This results in larger repositories and a higher number of designers contributing.

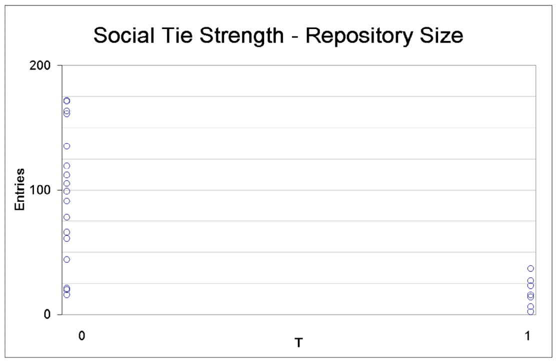

Fig. 5. Relation between tie strength $\boldsymbol{T}$ and repository size. Strong social ties generate smaller repositories, whilst weak ties produce larger and more variable repositories.

These results can be formulated as a principle of gatekeeping and social ties: Social groups where individuals have stronger links produce more stable gatekeeping, i.e. the process of selecting artifacts for a collective repository remains in the same hands for long periods of time. One direct result is that such repositories are of limited size than in equivalent societies where social ties are weaker. The artifacts of more designers that operate within weaker social spaces are more likely to be recognized by experts of the field.

3.2 Gatekeeping Rate ($\boldsymbol{G}$)

In this experiment the situational factor under inspection is given by the frequency of the gatekeeping role. Gatekeepers periodically select artifacts to enter the collective repository of their population. This nomination process is executed with a fixed rate during a simulation run. This is defined as the gatekeeping rate $\boldsymbol{G}$. When $\boldsymbol{G} = 100$, gatekeeping is executed sporadically at every 100 adoption steps. When $\boldsymbol{G} = 10$, gatekeeping takes place more often at every 10 adoption steps. The assumption is that in different design fields there may be different gatekeeping rates $\boldsymbol{G}$ when compared to the rate of adoption.

As may be expected, increasing gatekeeping activity generates larger repositories as shown in Fig. 6. The reason could be that when gatekeepers consider entries to the domain more often, there is a higher probability that more artifacts are selected. Perhaps more surprisingly, the relationship between $\boldsymbol{G}$ and the mean size of repositories is non-linear. In other words, a small decrease in the rate of frequent gatekeeping rapidly generates smaller repositories. However, when gatekeeping is sporadic, rate changes do not have a significant impact on the number of selections. Further, no significant correlation is observed between $\boldsymbol{G}$ and the mean quality of repositories entries. The latter is measured by the score assigned by gatekeepers to selected artifacts. This demonstrates that the frequency of selection does not determine the content of the repository. One possible explanation for the independence of these variables is that whilst more frequent gatekeeping includes entries with small improvements, more periodic gatekeeping includes artifacts with equivalent scores in larger increments. In other words, the smaller repositories generated by less frequent gatekeeping are more refined as they contain less entries but a similar mean quality.

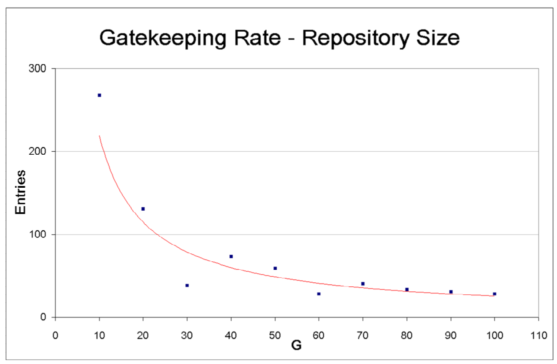

Fig. 6. Relation between gatekeeping rate $\boldsymbol{G}$ and the size of the domain. The shape of the relationship shows that small changes of frequent gatekeeping have a significant effect on the number of selected artifacts. In contrast, changes of more periodic gatekeeping have a marginal effect.

These results can be stated in a principle of gatekeeping frequency and domains: In fields where gatekeeping takes place frequently, larger domains are expected. When gatekeepers only select artifacts sporadically, the number of selections in the domain is smaller. However, these variations do not determine the mean quality assigned to domain entries. Therefore, sporadic gatekeeping is more efficient since same-quality levels are reached by fewer entries.

3.3 Population Size ($\boldsymbol{P}$)

In this experiment the effects of population size in the design and adoption of artifacts are considered. Population size $\boldsymbol{P}$ is given by the number of adopter agents in a population. $\boldsymbol{P}$ is assigned at initial time and remains fixed during a simulation run. The results presented here refer to cases where $\boldsymbol{P}$ is manipulated from $\boldsymbol{P} = 10$ to $\boldsymbol{P} = 100$. In principle larger $\boldsymbol{P}$ values can be studied but data collection of Monte Carlo runs becomes more arduous. In these experiments all other variables remain unchanged including the number of designers, the strength of social ties, and design and gatekeeping rates.

Unsurprisingly, increasing the number of adopters in a population generates a linear increase of the mean size of their repositories as shown in Fig. 7. This is a consequence of the definition of opinion leaders in our framework based on standard deviations from the mean. The number of opinion leaders increases linearly with $\boldsymbol{P}$, i.e. a population of 100 adopters has, on average, 10 times more gatekeepers than a population of 10 adopters. Further experimentation could be carried where the ratio of gatekeepers varies in order to study the effects of population size with constant number of gatekeepers.

As Fig. 8 shows, the quality of repository entries only varies marginally with $\boldsymbol{P}$. Figure 8a plots the mean repository scores attributed by gatekeepers Fig. 8b plots an independent measure of complexity of the entries. Both measures show that qualitatively, the selected artifacts are nearly equivalent.

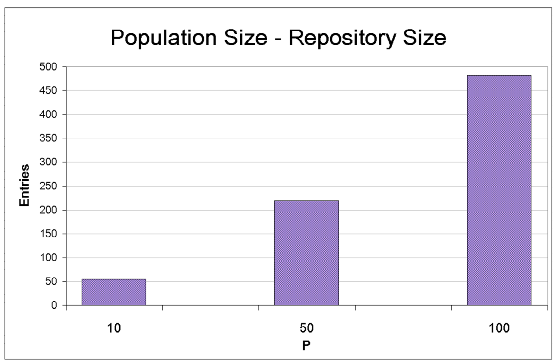

Fig. 7. Relation between population size $\boldsymbol{P}$ and the mean size of repositories. Since the number of gatekeepers is a ratio of the population, large populations generate proportionally large repositories.

These results suggest that whilst larger populations may generate proportionally larger domains, the mean quality of a large domain need not be significantly higher than that of less populated societies. In other words, when large social groups involve more gatekeeping, they tend to generate larger domains. However, large domains need not be of proportional quality. Small adopter societies can produce small repositories of comparable mean quality.

4 Discussion

This paper demonstrates a series of basic principles using a computational framework of design as a social activity based on the DIFI framework of creativity [6]. The results shown here demonstrate ways in which gatekeeping can play a fundamental role in the determination of creativity.

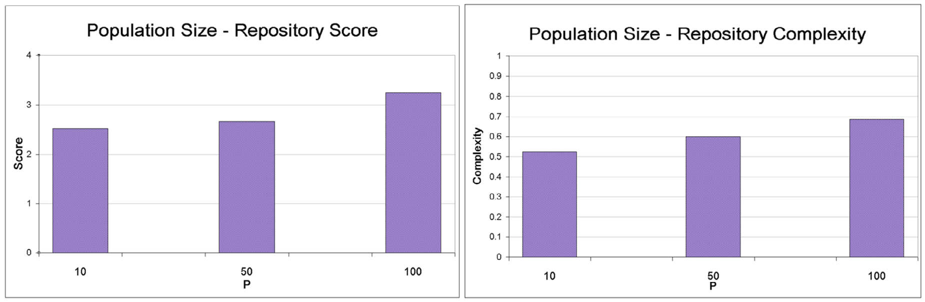

Fig. 8. (a) Relation between population size $\boldsymbol{P}$ and the mean score of repository entries and (b) Relation between population size $\boldsymbol{P}$ and the mean complexity of repository entries. Whilst large populations may generate large repositories, the quality of their entries is relatively similar.

Firstly, the nature and frequency of social relationships where opinions are exchanged could be a significant factor in the emergence of creative figures. The type and strength of the exchange of opinions may be responsible for the size and content of the domain. Secondly, the scheduling of expert assessments may also be a source of variations in the domain although it need not be related to the quality of its content. Lastly, the size of a social group may cause significant variations in the size of the domain but only marginal variations in its quality.

Whilst it is implausible to generalize these findings outside the scope of this rather simplistic framework, they point to similar conclusions from evidence in the literature. Gardner [7] concludes after studying the cases of Picasso and Freud that "in more hierarchical fields a small number of creators gain prominence and influence". An explanation is given by our finding that in fields where influence concentrates in a small and stable group of gatekeepers, domain entries are selected from a small number of designers. In contrast, when influence in the field is distributed, more variation of selection views causes larger domains with entries from a larger number of designers. Our experiments further suggest that this function could be nonlinear making it relatively easy to distribute influence in hierarchical fields.

Other aspects of gatekeeping need to be compared to evidence from creative cases. The aim of this paper has been to demonstrate the types of situational factors that are normally considered as exogenous processes in the study and modeling of creativity. Paradoxes of individual instabilities involving creative output and creative peak age may be explained by field characteristics demonstrating that creative individuals and algorithms are, by definition, not possible to be modeled in a social void.

Acknowledgements

This research is supported by an International Postgraduate Research Scholarship. Computing resources provided by Key Centre of Design Computing and Cognition.

References

1. Boden, M. (ed.): 1994, Dimensions of Creativity, MIT Press, Cambridge.
2. Amabile, TM.: 1983, The Social Psychology of Creativity, Springer-Verlag, New York.
3. Simonton, DK.: 1997, Creative productivity: a predictive and exploratory model of career trajectories and landmarks, Psychological Review, **104**(1), 66-89
4. Rinaldi, S., Cordone, R. and Casagrandi, R.: 2000, Instabilities in creative professions: a minimal model, Nonlinear Dynamics, Psychology, and Life Sciences, **4**(3), 255-273
5. Csikszentmihalyi, M.: 1997, Creativity, Flow and the Psychology of Discovery and Invention, HarperCollins, New York.
6. Feldman, DH., Csikszentmihalyi, M., and Gardner, H.: 1994, Changing the World, A Framework for the Study of Creativity, Praeger, Westport.
7. Gardner, H.: 1993, Creating Minds, an Anatomy of Creativity Seen Through the Lives of Freud, Einstein, Picasso, Stravinsky, Eliot, Graham and Gandhi, Basic Books, New York.
8. Cropley, A.: 1999, Definitions of creativity. In Runco, M. and Pritzker, S. (eds.), Encyclopedia of Creativity, Academic Press, San Diego, pp.511-524
9. Rubin, F.: 1974, A search procedure for Hamilton paths and circuits, Journal of the ACM, **21**, 576-580
10. Castelfranchi, C.: 2001, The theory of social functions: challenges for computational social science and multi-agent learning, Journal of Cognitive Systems Research, **2**(1), 5 -38.
11. Sosa, R. and Gero JS.: 2004, Diffusion of design ideas: gatekeeping effects. In Soo Lee, H. and Won Choi, J. (eds.), CAADRIA 2004, Yonsei University Press, Seoul.
12. Wasserman, S. and Faust, K.: 1994, Social Network Analysis: Methods and Applications, Cambridge University Press, Cambridge.
13. Eysenck, HJ. and Eysenck MW.: 1985, Personality and Individual Differences: A Natural Science Approach, Plenum Press, New York.
14. Granovetter, MS.: 1978, Threshold models of collective behavior, American Journal of Sociology, **83**(6), 1420-1443.
15. Liggett, TM.: 1985, Interacting Particle Systems, Springer-Verlag, New York.
16. Runco, M. and Pritzker, S. (eds.): 1999, Encyclopedia of Creativity, Academic Press, San Diego.
17. Granovetter, MS.: 1973, The strength of weak ties, American Journal of Sociology, **78**(6), 1360-1380.
18. Marsden, PV. and Campbell, K.: 1984, Measuring tie strength, Social Forces **63**(2): 482-501.

Evolving Creativity

Penousal Machado[1,2], Jorge Tavares[1], Amílcar Cardoso[1], Francisco B. Pereira[1,2], and Ernesto Costa[1]

[1] Instituto Superior de Engenharia de Coimbra, Quinta da Nora, 3030 Coimbra, Portugal
[2] Centre for Informatics and Systems of the University of Coimbra, Pólo II - Pinhal de Marrocos, 3030 Coimbra, Portugal
{machado, jast, amilcar, xico, ernesto}@dei.uc.pt

Abstract. Transformational creativity requires a change of the search space. As such, Evolutionary Computation (EC) approaches are incapable of transformational creativity. In this paper, we discuss how canonical EC techniques can be extended in order to yield the potential of transformational creativity. We present a formalized description of how this can be attained, and the experimental results achieved with a meta-evolutionary scheme.

1 Introduction

In her works on creativity, Margaret Boden (e.g. [1]) identifies two types of creativity: exploratory and transformational (for short: *e-creativity* and *t-creativity*), and considers the latest as more important. A typical EC approach is thought to yield the potential for *e-creativity*. However, it is deemed as being incapable of *t-creativity*. In this paper we propose several extensions of EC, which we consider to confer to EC the potential for *t-creativity*.

In section 2 we make a synthesis of key concepts related with creativity, and of a formalization of some of Boden's ideas. In the next section we present a generic EC algorithm, and propose the evolution of some of its components. Then, in section 4, we make a short overview of EC systems in which some of these items are evolved. In section 5 we establish a relation between the evolution of EC components and *t-creativity*. The experimental results attained with a meta-evolutionary approach are presented and discussed in section 6. Finally, we draw some conclusions and present future research directions.

2 Exploratory and Transformational Creativity

In this section we introduce some key concepts related with creativity that are relevant for the remainder of the paper. We do not address most of the questions in detail, since that is clearly beyond the scope of the paper.

Boden [1] establishes two axis for the characterization of creativity. The first relates with proprieties of the product, and establishes a distinction between Historical and Psychological creativity: creating something that was never created before versus creating something that is novel only to the eyes of the agent. This distinction, although important from a social perspective, is of little consequence for our study.

The second axis characterizes the type of creativity – exploratory versus transformational – and has wider implications. Boden views the creative process as a search of new objects in a conceptual space. This space can be extremely convoluted, and therefore some of its points hard to reach. The discovery one point of this type is deemed as *e-creativity*. While *e-creativity* is an exploration of a conceptual space, *t-creativity* refers to the change of the conceptual space itself. As such, great breakthroughs that provoke paradigm shifts fit into this class of creativity [1].

Wiggins [2] formalized some of the ideas presented by Boden. This formalization – which sheds a new light into the somewhat vague description provided by Boden, making it more clear and precise – will be described in the next section. For a more thorough description we refer the reader to the original paper.

2.1 Formalization

The formalization proposed by Wiggins [2] includes the following elements:

- U – The space of all possible concepts. Or, for parsimony, the set of all possible concepts relevant to a given domain.
- L – A language that enables the definition of constraints and construction rules.
- $[[.]]$ – An interpreter for selecting concepts from U according to a set of constraints specified in L.
- $\langle\langle . \rangle\rangle$ – A search engine for traversing U, or one of its subsets, according to rules specified in L.
- R – A set of rules, in L, that defines a subset of U.
- T – A set of rules, in L, defining the search strategy.
- E – A set of rules, in L, which allow the evaluation of concepts.

These elements allow the modelling of Boden's ideas with more precision. Lets represent by C the conceptual space that experts in a given area usually consider in the search for new concepts. This space is a subset of U, defined by the rules in R.

$$C = [[R]](U) \tag{1}$$

Additionally, T is a set of rules encoding a search engine that allows the traversal of C, thus defining the connectivity between points of the space:

$$c_{i+1} = \langle\langle R \cup T \rangle\rangle(c_i) \tag{2}$$

E-creativity is modelled as an exploration of the search space C, using the search methodology specified by T, that leads to new points of C which are highly valued by E. One can add the additional constraint of these points being hard to reach, but this is not strictly necessary. Moreover, E can take into account aspects other than adequacy (e.g. novelty). According to the formalization, *t-creativity* can be achieved in several ways, namely:

– By changing the set of rules, R, that defines the elements of C, thus creating a new conceptual space, C'.
– By changing T, the rules governing the traversal of C. In this case the elements of C do not change, what is changed is the connectivity between the points of C.

This distinction only became clear due to the formalization [2]. Wiggins also notices that: "...a change in E opens up a whole new area of conceptual space, and possibly of universe, for consideration." [2]. In a human society a change in E is probably not easy to achieve. Nevertheless, we put forward the possibility of this being yet another way of achieving *t-creativity*.

Wiggins' formalization allows the definition of *t-creativity* in precise terms. Simply put, *t-creativity* is *e-creativity* at the meta-level. An exploratory creative system can be described by a sextuple:

$$\langle U, L, [[.]], \langle\langle . \rangle\rangle, R, T, E \rangle \tag{3}$$

T-creativity is the search for a new R or T or both. Since they are both expressed in L, L becomes the space of all possible concepts. A new language L_L is needed, to allow the definition of constraints and rules for traversing this space. We also need R_L, a set of constraints in language L_L that specifies the search space of *R's* and *T's* that will be considered, and T_L a set of rules in L_L for traversing that space. Additionally, we have the corresponding interpreters: $\widehat{[[.]]}$ and $\widehat{\langle\langle . \rangle\rangle}$. Finally we need an evaluation function E_L, also in L_L, that assesses the quality of the *R's* and *T's*. Thus, *t-creativity* can be described by the following sextuple:

$$\langle L, L_L, \widehat{[[.]]}, \widehat{\langle\langle . \rangle\rangle}, R_L, T_L, E_L \rangle \tag{4}$$

Hence the argument of *t-creativity* being *e-creativity* at the meta-level [2]. This also gives rise to the possibility of considering further meta-levels of creativity. There are, of course, practical implications, which include how to build: R_L, T_L, and E_L. We will return to these issues later.

3 Evolutionary Computation

Historically EC is divided into four families namely: Evolution Strategies (ES); Evolutionary Programming (EP); Genetic Algorithms (GA); and Genetic Programming (GP). In spite of their differences they can all be seen as particular instances of the Generic Evolutionary Algorithm (GEA) presented in figure 1.

The first step is the creation of a random set of genotypes, $G(0)$. These genotypes are converted to phenotypes through the application of a mapping function (*map*). In most cases there is not a clear distinction between genotype and phenotype, so this step is typically omitted. The next step consists on the evaluation of the individuals. This is performed at the phenotype level using a fitness function, *eval*. The main evolutionary cycle follows. A set of parents is selected, using *sel*, followed by the application of genetic operators, *op*, which

```
t ← 0
G(0)  ← generate_random(t)
P(0)  ← map(G(0))
F(0)  ← eval(P(0))
while stop criterion not met do
      G′(t)      ← sel(G(t), P(t), F(t))
      G″(t)      ← op(G′(t))
      G(t + 1)   ← gen(G(t), G″(t))
      P(t + 1)   ← map(G(t + 1))
      F(t + 1)   ← eval(P(t + 1))
      t ← t + 1
end while
return result
```

Fig. 1. Generic Evolutionary Algorithm.

yields a new set of genotypes, $G''(t)$. The next steps consist in the generation of the phenotypes and their evaluation. The evolutionary cycle continues until some termination criterion is met.

We are now in position to return to the original motivation for this paper: "How can transformational creativity be achieved by means of Evolutionary Computation?". Assuming that EC has the potential to achieve *e-creativity*, which seems to be the case; and that *t-creativity* is *e-creativity* at the meta-level, which follows from the work of Wiggins [2]; then meta-evolution should yield the potential to perform *t-creativity.* More precisely, what we propose evolving the following aspects of the evolutionary algorithm: mapping function, genetic operators, selection procedure, replacement strategy, and evaluation function.

4 Related Work

The area of Adaptive Evolutionary Computation (AEC) focuses on the evolution of specific parameters of EC algorithms. Angeline [3] makes a formal definition and classification of AEC, proposing three levels of adaptation: population-level, individual-level and component-level.

There are several AEC approaches that allow the dynamic resizing of the genotype, allowing its expansion and contraction according to environmental requirements. Angeline and Pollack [4] propose the co-evolution of: a high-level representational language suited to the environment; and of a dynamic GA where the genotype size varies. Also related to genotype-phenotype mapping, is the work of Altenberg [5] about the notion of "evolvability" - the ability of a population to produce variants fitter than previous existing ones. In [6, 7] Altenberg explores the concept of Genome Growth, a constructional selection method in which the degree of freedom of the representation is increased incrementally. This work is directly connected to the concepts presented by Dawkins in [8], where he clearly differentiates genetics, the study of the relationships between genotypes in successive generations, from embryology, the study of the relationships between genotype and phenotype in any one generation. This leads us to the concept of embryogeny, the process of growth that defines how a genotype is mapped onto a phenotype, and to the work of Bentley. In [9], the use of such growth processes within evolutionary systems is studied. Three main types of EC embryogenies are identified and explained: external, explicit and implicit. A comparative study between these different types, using an evolutionary design problem, is also presented.

The evolution of the genetic operators is the obvious next step. In [10], Teller describes how genetic operators can be evolved using PADO, a graph based GP system. In [11] GP is extended to a co-evolutionary model, allowing the co-evolution of candidate solutions and genetic operators. Another approach to the evolution of genetic operators is described in [12]. In this study, an additional level of recombination operators is introduced, which performs the recombination of a pool of operators. In [13] Spector and Robinson discuss how an autoconstructive evolutionary system can be attained using a language called Push.

5 Evolutionary Transformational Creativity

In this section we analyze the consequences of evolving several components of the traditional EC, and propose propose several alternatives for their evolution.

5.1 Transforming the Search Space

We start by analyzing what kind of transformation of the search space can be achieved by changing the *map* function. This function is responsible for the mapping between genotype and phenotype. A phenotype is a fully grown individual, or, from a more computational point of view, a candidate solution to a given problem. A genotype, is a piece of genetic code that once expressed via the *map* function, results in a phenotype. The genotype has no independent meaning, it only gains meaning after the application of a mapping function.

We can establish a connection between the *map* function and Wiggins' formalization. Considering that *map* is expressed in some language, that $[[.]]$ is an interpreter for that language, and that u is the set of all genotypes – we have:

$$c = [[map]](u), \tag{5}$$

where c is the space of all possible phenotypes under *map*. Thus, *map* is *roughly* equivalent to R.

To be totally equivalent we would have to consider the space of genotypes, u, to be equivalent to U. According to Wiggins' formalization C and U are both concept spaces, and C should be a subset of U. In general, these propositions do not hold for c and u. In the typical scenario u and c are sets of different types, genotypes and phenotypes. Thus, we can establish a parallelism between C and c, however, in general, the same parallelism cannot be established between U and u. Nevertheless, for particular cases, e.g. when there is an identity relation between genotypes and phenotypes, u can be considered equivalent to U. Moreover, establishing an equivalence between u and U is not strictly necessary. Since *t-creativity* is about the transformation of C (or T), the equivalence between C and c is, from our view point, sufficient to be in accordance with Wiggins' formalization.

Additionally, and due to the genetic operators being applied at the genotype level, a transformation of *map* may also lead to a change of the connectivity of the space, and thus of T.

The way the space is traversed depends, mostly on the operators employed. They are the main ingredient in the definition of the connectivity of the space. In a less direct way, the connectivity also depends on the selection procedure, and on the replacement strategy (*gen*). As such, a change of *op*, *sel*, or even *gen*, can be seen as a transformation of T – the rules governing the traversal of C.

Like we stated before, the genetic operations are performed at the genotype level. This causes no conflict with the formalization. Equation 2.1 clearly states that the interpreter $\langle\langle . \rangle\rangle$ is applied to $R \cup T$.

Changing *sel* or *gen* does not appear to be as interesting as changing the genetic operators. Nevertheless, a change of these functions has the potential to change the search method, and as such yield *t-creativity*.

The relation between the fitness function, *eval*, and E is obvious. In section 2.1 we suggested that a change in E could also be a type of *t-creativity*. In the case of a EC approach this is certainly the case. The *sel* of the progenitors takes into account their fitness value. Therefore, a change in E can change the way the space is traversed.

5.2 Evolving EC Components

The most obvious approach to the evolution of EC components is the use of Meta-Evolution. Lets assume we are interested in evolving one of the components of EC, for instance the mapping function, and that we resort to GP to evolve populations of these functions.

Each genotype, G_i^{map}, is an encoding of a candidate mapping function; once expressed, via map^{map}, it results in a phenotype, P_i^{map}. Thus, the phenotypes are mapping functions expressed in some language L^{map}. For simplicity sake, we can assume that this language to be Turing Complete, which implies that any computable mapping function can be evolved.

We also need to define: sel^{map}, op^{map} and gen^{map}. Like map^{map} and $eval^{map}$ these functions are static. Therefore, we can use standard GP selection, operators and replacement strategy.

Moreover, we need to develop a way to assign fitness to different mapping functions, $eval^{map}$. One possibility is to use a set of rules that specify the characteristics that are considered desirable in a mapping function, and assign fitness based on the accordance with those rules. Although technically possible, this can be both difficult and pointless. To attain *t-creativity* we must be *e-creative* at this level. The chances of finding an interesting, innovative, and adequate mapping function, with a search procedure guided by a pre-established set of rules, or heuristics, seems slim. Moreover, in most scenarios, we might not even have a good idea about the kind of mapping function we are interested in. Additionally, EC is usually good at finding holes in the fitness function.

There is, however, something that we usually can take for granted: we are mainly interested in mapping functions that promote the discovery of good solutions for the original problem. We can, for each individual being evaluated, run an EC algorithm in which P_i^{map} is used as mapping function. By changing the GEA presented in figure 1 so that we can pass as arguments one, or several, of the following functions: *map*, *sel*, *op*, *gen* and *eval*, we can use an $eval^{map}$ such as the one presented in figure 2. The fitness of each phenotype is the result of a lower level EC. This result can indicate, for instance: the fitness of the best individual (at the lower level); the time necessary to reach the optimum; the average fitness of the last population; etc.

$eval^{map}(P^{map})$
  for $i = 1$ **to** $\#(P^{map})$ **do**
    $F_i^{map} \leftarrow AEG(P_i^{map})$
  endfor
  return F^{map}

Fig. 2. Meta-level fitness function.

We have stated that the evolved mapping functions could be expressed in some Turing Complete language. If this is the case, since the lower level EC runs P_i^{map}, we must deal with the Halting Problem. The typical solution is to impose a time constraint. Thus, if P_i^{map} does not halt after a pre-specified amount of time it is assumed that it will never stop and, accordingly, its fitness value will be low.

It should be more or less obvious that we can employ the exact same strategy to evolve: *sel*, *op*, *gen* or *eval*. If we are evolving evaluation functions, the lower level EC is guided by P_i^{eval}. However, we are interested in a P_i^{eval} which allows the discovery of individuals which are fit accordingly to some original fitness function. As such, the return value of GEA should reflect its performance according to this original function.

There is, of course, the possibility of adding more levels enabling the evolution of several components. Alternatively, we can also evolve several components simultaneously. In this case, each genotype would be an encoding of several functions, and the phenotype a set of functions, which are passed to the lower level EC. For instance, considering that all components are being evolved, we would have $P_i^{\{map,sel,op,gen,eval\}}$.

The main problem of the architecture presented in this section is its high computational cost. There are several other alternatives to the evolution of EC components, among which: Dual-Evolution and Co-Evolution. Due to space restrictions these approaches will not be analyzed in this paper, for a brief overview please consult [15].

6 Experimental Results

To test our ideas we decided to apply meta-evolution to evolve mapping functions. We use a two level meta-evolution scheme composed by a GP algorithm and a GA. At the higher level we have the GP algorithm, which is used to evolve the mapping functions. At the lower level we have the GA, whose task is finding the maximum value of a mathematical function. The goal is to evolve mappings that help the GA to accomplish its task.

The function being optimized by the GA, $f(x)$, is defined over the interval $[0, 1]$, and is the sum of $f_{peak}(x)$ and $f_{wave}(x)$, which are specified as follows:

$$f_{peak}(x) = max(0,\ |1 - 2|x - peak| - (1 - \frac{1}{r})| \times 0.1 \times r) \qquad (6)$$

$$f_{wave}(x) = cos(2\pi \times r \times (x - peak)) \qquad (7)$$

f_{wave} creates a sine wave with r repetitions in the $[0, 1]$ interval, returning values between -1 and 1. By adding f_{peak} we change the values of one of the repetitions, making it reach a maximum value of 1.1. In figure 3 we present a graph of $f(x)$. To increase the complexity of the problem, variable r is set to 100, which means that the wave function repeats itself 100 times in the $[0, 1]$ interval.

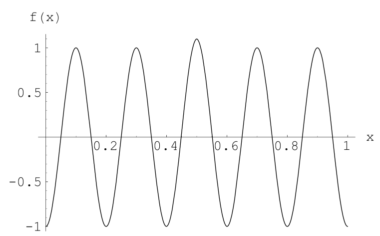

Fig. 3. Test function, $r = 5$, $peak = 0.5$

By changing the value of the variable *peak* we can change the coordinate of the maximum of $f(x)$. This ability to change the maximum is fundamental. If this value does not change the GP algorithm could find programs that output a constant x value corresponding to the maximum of $f(x)$. This is not, obviously, the desired behavior. What we aim to achieve is a GP program that transforms the search space in a way that helps the GA to find the maximum value. Thus, for each value of x the GP programs should compute a new value, x'. The re-organization of the search space induced by the x to x' transformation should make the task of finding the optimum easier.

To ensure that it is possible to find a good mapping function we decided to design one by hand. We were able to develop the following function:

$$g_{optimal}(x) = \frac{x + floor(frac(x \times 10000) \times r)}{r} \quad (8)$$

Where $frac$ returns the fractional part. This function has the effect of folding the space r times, and then expanding it back to $[0, 1]$. By using this mapping function the topology of the search space is changed, resulting in a less convoluted fitness landscape. In figure 4 we present a chart of this mapping function, $g_{optimal}(x)$, and of the search space resulting from its application, $f(g_{optimal}(x))$. Since $g_{optimal}$ maps $[0, 1]$ to $[0, 1]$, R is not changed. Instead, what is changed is the connectivity of the points of the space and, as such, T. This becomes possible due to the genetic operations being performed at the genotype level.

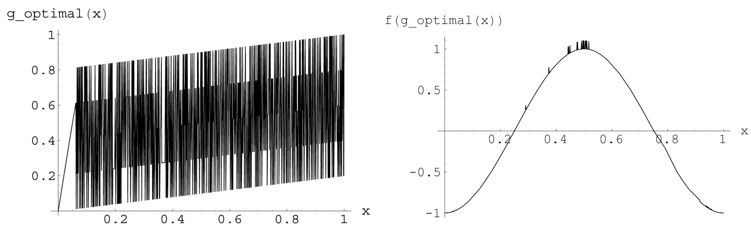

Fig. 4. On the left $g_{optimal}(x)$; on the right $f(g_{optimal}(x))$, $r = 5$.

To assign fitness, we run, for each GP individual, $mapping_j$ a GA. Each GA genotype, G_i^{sol}, is a binary string of size 50, encoding a value, $G_i'^{sol}$, in the $[0, 1]$ interval. The GP individual is used as mapping function for the GA. The value $G_i'^{sol}$ will be mapped to x'_i (thus, $x'_i = mapping_j(G_i'^{sol})$). x'_i is then used as the x coordinate for the function being optimized by the GA, f.

In order to get a good estimate of the quality of the mapping $functions$ we perform 30 runs of the GA for each GP individual. To prevent the specialization of the GP individuals, the value of $peak$ is randomly chosen at the beginning of each GA run. The fitness the GP individual is equal to the average fitness of the best individual of the last population of the lower level GA.

We use the following function and terminal sets: $\{+, -, \%, \times, floor, frac\}$, where % is the protected division; $= \{G_i'^{sol}, 1, 10, 100\}$, where $G_i'^{sol}$ is a variable holding the value of the GA genotype that is currently being mapped.

The settings for the GP algorithm were the following: Population size = 100; Number of generations 500; Swap-Tree crossover; Generate Random Tree mutation; Crossover probability = 70%; Mutation probability=20%; Maximum tree depth = 10. The settings for the GA where the following: Population size = 100; Number of generations = 30,100; Two point crossover; Swap Mutation; Crossover probability = 70%; Mutation probability=$\{1\%, 2.5\%, 5\%\}$.

6.1 Analysis of the Results

The main objective of our approach is to find a mapping function that consistently improves the performance of the GA algorithm. To evaluate the experimental results we need some reference points. Therefore, we conducted a series of experiments in which the mapping function was not subjected to evolution. In these tests we used the following static mapping functions: $g_{optimal}$, already described; and $g_{identity}$ with $g_{identity}(x) = G'^{sol}$. These results will be compared with the ones obtained using as mapping functions the best individuals of each GP run, $g_{evolved}$.

Table 1 shows the average fitness of the best individual of the last population of a traditional GA (Number of generations = 100) using as mapping functions $g_{identity}$, $g_{optimal}$ and $g_{evolved}$. The results are averages of 100 runs for the static mappings and of 3000 runs for the evolved mappings (100 runs per each evolved mapping).

Table 1. Average fitness of the best individual of the last population. The entries in bold indicate a statistically significant difference between these values and the corresponding $g_{identity}$ values ($\alpha = 0.01$).

Mutation	$g_{identity}$	$g_{optimal}$	$g_{evolved30}$	$g_{evolved100}$
1%	1.0134617	**1.0806543**	**1.0647961**	**1.0632421**
2.5%	1.0242724	**1.0942163**	**1.0843632**	**1.0821193**
5%	1.0323492	**1.0982930**	**1.0993311**	**1.0946481**

As expected using $g_{optimal}$ considerably improves the performance of the algorithm, yielding averages close to the maximum attainable value, 1.1. Table 1 shows that the use of the evolved mapping functions significantly improves the performance of the GA. However, the results attained are usually inferior to the ones achieved using $g_{optimal}$. This difference diminishes as the mutation rate increases. Additionally, using a smaller number of generations in the lower level GA leads to better overall results. The higher evolutionary pressure increases the need for the transformation of the search space.

To get a better grasp of how the use of the evolved mapping functions alter the GA, we present, in figure 5, the evolution of the fitness of the best individual during the GA run. For static mappings the fitness increases abruptly in the first generations, stagnating for the remainder of the run; with the evolved mappings the fitness increases steadily during the entire run. An analysis of the evolution of the average fitness of the GA populations gives insight to how the evolved mappings are improving the GA performance. The use of evolved mappings decreases significantly the average fitness of the populations. These results indicate that the evolved mappings improve the performance of the GA by promoting phenotypic diversity, preventing the early stagnation of the GA runs.

The evolved mappings are not similar to $g_{optimal}$, which cannot be considered surprising. As is often the case when analyzing the results of a GP program, it

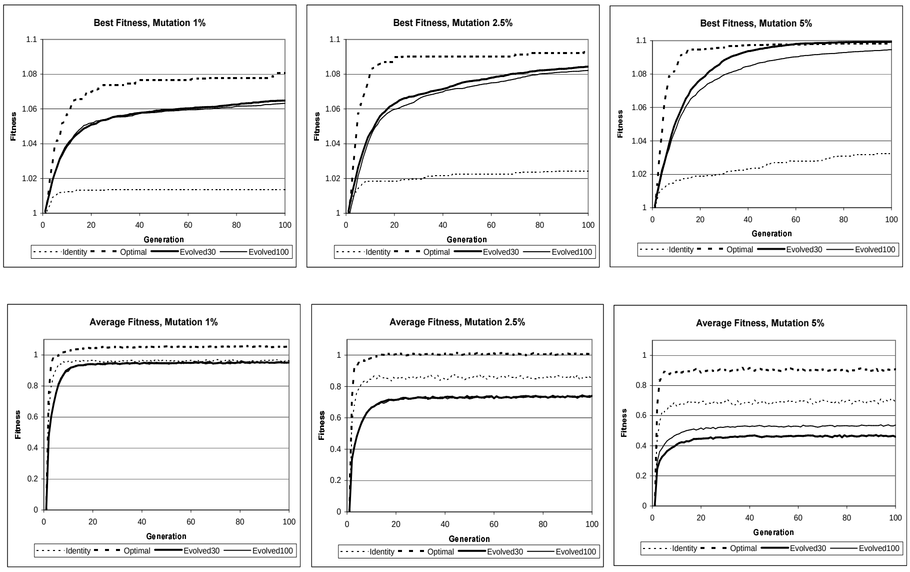

Fig. 5. Evolution of the fitness of the best individual and of the average fitness of the GA populations.

is not clear how the evolved mappings solve the problem of improving the GA performance. The charts suggest that reducing the number of GA generations used in the GP fitness assignment procedure, thus increasing the difficulty of the evolved mappings task, leads to better results. Taking into account the average of the GA populations when assigning a fitness value for the GP individuals, may also prove useful to achieve mappings closer to $g_{optimal}$.

7 Conclusions and Further Work

In this paper we discussed how the canonical EC algorithm can be extended in order to yield the potential for *t-creativity*. The proposed changes involve the evolution of several components of EC which are typically static. We establish a relation between the evolution of these components and the change of R and T, introduced in Wiggins' formalization of *t-creativity* [2]. Additionally, the evolution of these components may prove useful in improving the EC performance, lessen the burden of researchers, and provide indications about the characteristics of the problems being solve.

The attained results are promising, and provide pointers for the improvement of the approach. Future research will include: making a wider set of experiments; applying the proposed approach to a different set of domains; and using dual-evolution and co-evolution to evolve EC components.

References

1. Boden, M.A.: The Creative Mind: Myths and Mechanisms. Basic Books (1990)
2. Wiggins, G.: Towards a more precise characterisation of creativity in ai. In: Proc. of the ICCBR 2001 Workshop on Creative Systems, Vancouver, Canada (2001)
3. Angeline, P.J.: Adaptive and self-adaptive evolutionary computations. In: Computational Intelligence: A Dynamic Systems Perspective. IEEE Press (1995)
4. Angeline, P.J., Pollack, J.B.: Coevolving high-level representations. In: Artificial Life III. Volume XVII of SFI Studies in the Sciences of Complexity. (1994) 55–71
5. Altenberg, L.: The evolution of evolvability in genetic programming. In Kinnear, K.E., ed.: Advances in Genetic Programming. MIT Press (1994) 47–74
6. Altenberg, L.: Evolving better representations through selective genome growth. In: Proceedings of the 1st IEEE Conference on Evolutionary Computation. Part 1 (of 2), Piscataway N.J., IEEE (1994) 182–187
7. Altenberg, L.: Genome growth and the evolution of the genotype-phenotype map. In Banzhaf, W., Eeckman, F.H., eds.: Evolution as a Computational Process. Springer-Verlag, Berlin (1995) 205–259
8. Dawkins, R.: The evolution of evolvability. In: Artificial Life, SFI Studies in the Sciences of Complexity. Volume VI., Addison-Wesley (1989) 201–220
9. Bentley, P., Kumar, S.: Three ways to grow designs: A comparison of embryogenies for an evolutionary design problem. In: Proceedings of the Genetic and Evolutionary Computation Conference. Volume 1. (1999) 35–43
10. Teller, A.: Evolving programmers: The co-evolution of intelligent recombination operators. In Angeline, P.J., Kinnear, Jr., K.E., eds.: Advances in Genetic Programming 2. MIT Press, Cambridge, MA, USA (1996) 45–68
11. Edmonds, B.: Meta-genetic programming: Co-evolving the operators of variation. CPM Report 98-32, Centre for Policy Modelling, Manchester Metropolitan University, UK, Aytoun St., Manchester, M1 3GH. UK (1998)
12. Kantschik, W., Dittrich, P., Brameier, M., Banzhaf, W.: Meta-evolution in graph GP. In: Genetic Programming: Second European Workshop, Springer (1999) 15–28
13. Spector, L., Robinson, A.: Genetic programming and autoconstructive evolution with the push programming language. Genetic Programming and Evolvable Machines **3** (2002) 7–40
14. Koza, J.R.: Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems. Technical Report STAN-CS-90-1314, Stanford University (1990)
15. Tavares, J., Machado, P., Cardoso, A., Pereira, F.B., Costa, E.: On the evolution of evolutionary algorithms. In Keijzer, M., O'Reilly, U.M., Lucas, S.M., Costa, E., Soule, T., eds.: Genetic Programming 7th European Conference, EuroGP 2004, Proceedings. Volume 3003 of LNCS., Coimbra, Portugal, Springer-Verlag (2004) 389–398

Computational Abduction
The Extra-Theoretical Dimension of Scientific Creativity

Lorenzo Magnani

Department of Philosophy and Computational Philosophy Laboratory,
University of Pavia, 27100 Pavia, Italy,
lmagnani@unipv.it

Abstract Classical philosophers have offered a number of ways of describing hypotheses generation, but often they aimed at demonstrating that the activity of generating hypotheses is paradoxical, illusory or obscure, and thus not analyzable. Those descriptions are often so far from Peircian pragmatic prescription and so abstract to result completely unknowable and obscure. The "computational turn" gives us a new way to understand creative processes in a strictly pragmatic sense. In fact, by exploiting artificial intelligence and cognitive science tools, computational philosophy allows us to test concepts and ideas previously conceived only in abstract terms. It is in the perspective of these *actual computational models* that I find the central role of *abduction* in the explanation of creative reasoning in science. Creativity and discovery are no more seen as a mysterious non rational or irrational process, but, thanks to constructive accounts, as a complex relationship among different inferential steps that can be clearly analyzed and identified. I maintain that the computational philosophy analysis of *model-based* and *manipulative* abduction and of *external* and *epistemic mediators* is important not only to delineate the actual practice of abduction, but also to further enhance the development of programs computationally adequate in rediscovering, or discovering for the first time, for example, scientific hypotheses or mathematical theorems.

1 Computational Modeling and the Inexplicability of Creativity

Creativity is certainly an important aspect of our definition of "intelligence" but the literature associates many different notions to creativity. This ambiguity has brought to a lack of consensus in the research community. The common views associate to creativity unusual and mysterious qualities that drive the concept of creativity to a confused verbosity. Statements like "to break the rules", "to think different", "to destroy one Gestalt in favor of a better one", and "to arrange old elements into a new form", present in the field of psychological research on creativity since 1950s, certainly do not clarify the topic, and seem to lead to the Freudian conclusion that creativity cannot be understood. This conclusion has also been supported by many philosophers of science who studied conceptual change in science during the second half of the last century. They distinguished between a logic of discovery and a logic of justification (i.e. between the psychological side of creation and the logic argument of proving new dis-covered ideas by facts). The consequent conclusion was that a logic of discovery (and a rational model of discovery) could not exist: scientific conceptual change is cataclysmic and sometimes irrational, dramatic, incomprehensible and discontinuous.[1] Many studies already argued that creativity can be understood (for example [2,3], but paid attention mainly to the psychological and experimental aspects, disregarding the philosophical, logical, and computation ones. In AI research, however, since Simon, two characteristics seem to be associated to creativity: the novelty of the product and the unconventionality of the process that leads to the new product [4]. I maintain we can overcome many of the difficulties of creativity studies developing a theory of abduction, in the light of Charles Sanders Peirce's first insights.

We have said that philosophers of science in the twentieth century, following the *revolutionary* theory developed by Kuhn [5], have traditionally distinguished between the logic of discovery and the logic of justification.[2] Most have concluded that no logic of discovery exists and, moreover, that a "rational" model of discovery is impossible. In short, scientific creative reasoning should be non-rational or irrational and there is no reasoning to hypotheses. The problem is that the definition of concepts like "creative" and "discovery" is *a priori*. Following Peirce, the definitions of concepts of this sort have not usually rested upon any observed facts, at least not in any great degree; even if sometimes these beliefs are in harmony with natural causes. They have been chiefly adopted because their fundamental propositions seemed "agreeable to reason". That is, we find ourselves inclined to believe them. Usually this frame leads to a proliferating verbosity, in which theories are often incomprehensible and bring to some foresight just by intuition. But a theory which needs intuition to determine what it predicts has poor explanatory power. It just "makes of inquiry something similar to the development of taste" [8, p. 254].

A suggestion that can help to solve the enigma of discovery and creativity comes from the "computational turn" developed in the last years. Taking advantage of modern tools of logic, artificial intelligence, and other cognitive sciences, computational philosophy is able to construct actual models of studied processes. It is an interesting constructive rational alternative that, disregarding the most abstract level of philosophical analysis, can offer clear and testable architectures of creative processes.

2 The Centrality of Abduction

The development of human society has now reached a technological level in which issues concerning the creation and dynamics of information - especially in science - are absolutely crucial. Gradually, philosophical methods and problems are studied and understood in terms of the new information-theoretic notions. Inside the computational philosophy framework, a new paradigm, aimed at unifying the different perspectives and providing some design insights for future ones, rises by emphasizing the significance of the concept of *abduction*, in order to illustrate the problem-solving process and to propose a unified and rational epistemological model of scientific discovery, diagnostic reasoning, and other kinds of creative reasoning [9]. The concept of abduction

[1] The Imre Lakatos' epistemology of "proofs and refutations" is certainly one exception [1].
[2] A perspective originally established by Reichenbach [6] and Popper [7].

nicely ties together both issues related to the dynamics of information and its systematic embodiment in segments of various types of knowledge.

Abduction is the process of *inferring* certain facts and/or laws and hypotheses that render some sentences plausible, that *explain* or *discover* some (eventually new) phenomenon or observation; it is the process of reasoning in which explanatory hypotheses are formed and evaluated. There are two main epistemological meanings of the word abduction [9]: 1) abduction that only generates "plausible" hypotheses ("selective" or "creative") and 2) abduction considered as inference "to the best explanation", which also evaluates hypotheses. An illustration from the field of medical knowledge is represented by the discovery of a new disease and the manifestations it causes which can be considered as the result of a creative abductive inference. Therefore, "creative" abduction deals with the whole field of the growth of scientific knowledge. This is irrelevant in medical diagnosis where instead the task is to "select" from an encyclopedia of prestored diagnostic entities. We can call both inferences ampliative, selective and creative, because in both cases the reasoning involved amplifies, or goes beyond, the information incorporated in the premises.

Theoretical abduction[3] certainly illustrates much of what is important in creative abductive reasoning, in humans and in computational programs, but fails to account for many cases of explanations occurring in science when the exploitation of environment is crucial. It fails to account for those cases in which there is a kind of "discovering through doing", cases in which new and still unexpressed information is codified by means of manipulations of some external objects (*epistemic mediators*, cf.. below in this paper). The concept of *manipulative abduction*[4] captures a large part of scientists thinking where the role of action is central, and where the features of this action are implicit and hard to be elicited: action can provide otherwise unavailable information that enables the agent to solve problems by starting and by performing a suitable abductive process of generation or selection of hypotheses.

2.1 Sentential models of abductive reasoning

Many attempts have been made to model abduction by developing some formal tools in order to illustrate its computational properties and the relationships with the different forms of deductive reasoning [11]. Some of the formal models of abductive reasoning are based on the theory of the *epistemic state* of an agent [12], where the epistemic state of an individual is modeled as a consistent set of beliefs that can change by expansion and contraction (*belief revision framework*). I call sentential kinds of logical models [9].[5]

---

[3] Magnani [9] introduces the concept of theoretical abduction as a form of internal processing. He maintains that there are two kinds of theoretical abduction, "sentential" (formal, see below), related to logic and to verbal/symbolic inferences, and "model-based", related to the exploitation of models such as diagrams, pictures, etc, cf. below in this paper.

[4] Manipulative abduction and epistemic mediators are introduced and illustrated in [10] and [9].

[5] I have discussed in [13] the nature of the kinds of inconsistencies captured by these formalisms and shown how they do not adequately account for some roles played by anomalies, conflicts, and contradictions in many forms of explanatory reasoning.

Deductive models of abduction may be characterized as follows: an explanation for β relative to background theory T will be any α that, together with T, entails β (normally with the additional condition that $\{\alpha\} \cup T$ be consistent). Such theories are usually generalized in many directions: first of all by showing that explanations entail their conclusions only in a *defeasible* way (there are many potential explanations), so joining the whole area of so-called nonmonotonic logic or of probabilistic treatments; second, trying to show how some of the explanations are relatively implausible, elaborating suitable technical tools (for example in terms of modal logic) able to capture the notion of preference among explanations. Hence, we may require that an explanation makes the observation simply sufficiently probable [14] or that the explanations that are more likely will be the "preferred" explanations: the involvement of a cat in breaking the glass is less probable than the effect of wind. Finally, the deductive model of abduction does not authorize us to explain facts that are inconsistent with the background theory notwithstanding the fact that these explanations are very important and ubiquitous, for instance in diagnostic applications, where the facts to be explained contradict the expectation that the system involved is working according to specification.

In [12] Boutilier and Becher provide a formal account of the whole question in term of belief revision: if believing A is sufficient to induce belief in B, then A (epistemically) *explains* B; the situation can be semantically illustrated in terms of an ordering of plausibility or normality which is able to represent the epistemic state of an agent. The conflicting observations will require explanations that compel the agent to withdraw its beliefs (hypotheses), and the derived conditional logic is able to account for explanations of facts that *conflict* with the existing beliefs. The authors are able to reconstruct, within their framework, the two main paradigms of model-based diagnosis, and consistency-based diagnosis providing an alternative semantics for both in terms of a plausibility ordering over possible worlds.

These sentential models of abduction exclusively deal with selective abduction (diagnostic reasoning)[6] and relate to the idea of preserving *consistency*. Exclusively considering the sentential view of abduction does not enable us to say much about creative processes in science, and, therefore, about the nomological and most interesting creative aspects of abduction. It mainly refers to the *selective* (diagnostic) and merely *explanatory* aspects of reasoning and to the idea that abduction is mainly an inference *to the best explanation* [9].[7]

2.2 Model-based abduction and its external dimension

Computational philosophy taught us how to provide a suitable framework for constructing actual models of the most interesting cases of conceptual changes in science: we do not have to limit ourselves to the *sentential* view of theoretical abduction, mainly related to the "internal", mental aspect of reasoning, but we have to consider a broader *inferential* one: the *model-based* sides of creative abduction (cf. below).

[6] As previously indicated, it is important to distinguish between *selective* (abduction that merely selects from an encyclopedia of pre-stored hypotheses), and *creative* abduction (abduction that generates new hypotheses).

[7] For more details on the recent sentential models of abduction cf. the recent book [15].

From Peirce's philosophical point of view, all thinking is in signs, and signs can be icons, indices or symbols. Moreover, all inference is a form of sign activity, where the word sign includes "feeling, image, conception, and other representation" [16, 5.283], and, in Kantian words, all synthetic forms of cognition. That is, a considerable part of the thinking activity is model-based. Of course model-based reasoning acquires its peculiar creative relevance when embedded in abductive processes, so that we can individuate a *model-based abduction*. Hence we must think in terms of *model-based abduction* (and not in terms of sentential abduction) to explain complex processes like scientific conceptual change. Different varieties of *model-based abductions* [17] are related to the high-level types of scientific conceptual change [see, for instance, [18]].

Following Nersessian [19,20], the term "model-based reasoning" is used to indicate the construction and manipulation of various kinds of representations, not mainly sentential and/or formal, but mental and/or related to external mediators, such as visualizations, simulations, exploitation of artifacts, etc.

Manipulative abduction [9] - contrasted with theoretical abduction - happens when we are thinking through doing and not only, in a pragmatic sense, about doing. So the idea of manipulative abduction goes beyond the well-known role of experiments as capable of forming new scientific laws by means of the results (nature's answers to the investigator's question) they present, or of merely playing a predictive role (in confirmation and in falsification). Manipulative abduction refers to an extra-theoretical behavior that aims at creating communicable accounts of new experiences to integrate them into previously existing systems of experimental and linguistic (theoretical) practices. The existence of this kind of extra-theoretical cognitive behavior is also testified by the many everyday situations in which humans are perfectly able to perform very efficacious (and habitual) tasks without the immediate possibility of realizing their conceptual explanation [21]. We can find a similar situation also in the process of scientific creativity. Too often, in the cognitive view of science, it has been underlined that conceptual change just involves a *theoretical* and "internal" replacement of the main concepts. But usually researchers forget that a large part of this processes are instead due to *practical* and "external" *manipulations* of some kind, prerequisite to the subsequent work of theoretical arrangement and knowledge creation. When these processes are creative we can speak of manipulative abduction (cf. above). Scientists sometimes need a first "rough" and concrete experience of the world to develop their systems, as a *cognitive-historical* analysis of scientific change [22,23] has carefully shown.

3 Knowledge Creation and External Mediators

Even if, of course, a large portion of the complex environment of a thinking agent is internal, and consists of the proper software composed of the knowledge base and of the inferential expertise of the individual, nevertheless a "real" cognitive system is composed by a "distributed cognition" among people and some "external" objects and technical artifacts [21,24].

For example, in the case of the construction and examination of diagrams in geometrical reasoning, specific experiments serve as states and the implied operators are the manipulations and observations that transform one state into another. The geometrical outcome is dependent upon practices and specific sensory-motor activities performed on a non-symbolic object, which acts as a dedicated external representational medium supporting the various operators at work. There is a kind of an epistemic negotiation between the sensory framework of the geometer and the external reality of the diagram [25]. This process involves an external representation consisting of written symbols and figures that for example are manipulated "by hand". The cognitive system is not merely the mind-brain of the person performing the geometrical task, but the system consisting of the whole body (cognition is *embodied*) of the person plus the external physical representation. In geometrical discovery the whole activity of cognition is located in the system consisting of a human together with diagrams.

3.1 Manipulative Abduction, External Representations, Epistemic Mediators

Various *templates* of manipulative behavior exhibit some regularities. The activity of manipulating external things and representations is highly conjectural and not immediately explanatory: these templates are "hypotheses of behavior" (creative or already cognitively present in the scientist's mind-body system, and sometimes already applied) that abductively enable a kind of epistemic "doing". Hence, some templates of action and manipulation can be selected in the set of the ones available and pre-stored, others have to be created for the first time to perform the most interesting creative cognitive accomplishments of manipulative abduction.

Some common features of the tacit templates of manipulative abduction (cf. Figure 1), that enable us to manipulate things and experiments in science are related to: 1. sensibility towards the aspects of the phenomenon which can be regarded as *curious* or *anomalous*; manipulations have to be able to introduce potential inconsistencies in the received knowledge (Oersted's report of his experiment about electromagnetism is devoted to describing some anomalous aspects that did not depend on any particular theory of the nature of electricity and magnetism); 2. preliminary sensibility towards the *dynamical* character of the phenomenon, and not to entities and their properties, common aim of manipulations is to practically reorder the dynamic sequence of events into a static spatial one that should promote a subsequent bird's-eye view (narrative or visual-diagrammatic); 3. referral to experimental manipulations that exploit *artificial apparatus* to free new possible stable and repeatable sources of information about hidden knowledge and constraints (Davy set-up in term of an artifactual tower of needles showed that magnetization was related to orientation and does not require physical contact); 4. various contingent ways of epistemic acting: *looking* from different perspectives, *checking* the different information available, *comparing* subsequent events, *choosing*, *discarding*, *imaging* further manipulations, *re-ordering* and *changing relationships* in the world by implicitly *evaluating* the usefulness of a new order (for instance, to help memory).

The whole activity of manipulation is in fact devoted to building various external *epistemic mediators*[8] that function as an enormous new source of information and

[8] This expression, introduced by Magnani [9], is derived from the cognitive anthropologist Hutchins [21], who coined the expression "mediating structure" to refer to various external tools that can be built to cognitively help the activity of navigating in modern but also in

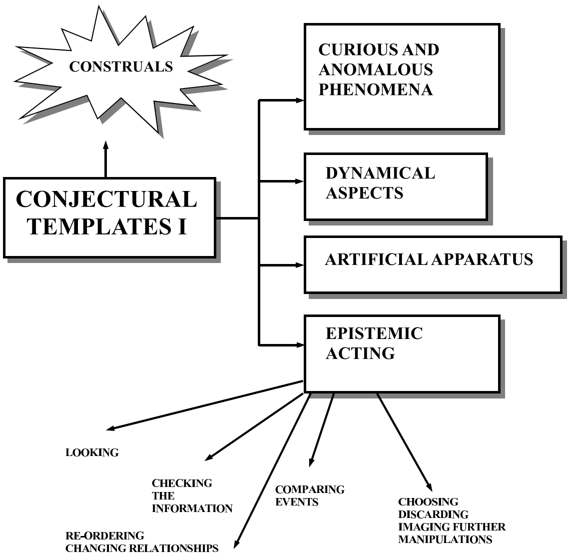

Figure 1. Conjectural templates I.

knowledge. Therefore, manipulative abduction represents a kind of redistribution of the epistemic and cognitive effort to manage objects and information that cannot be immediately represented or found internally (for example exploiting the resources of visual imagery).[9]

From the point of view of everyday situations manipulative abductive reasoning and epistemic mediators exhibit other very interesting templates (we can find the first three in geometrical constructions)(cf. Figure 2): 1. action elaborates a *simplification* of the reasoning task and a redistribution of effort across time [21], when we need to manipulate concrete things in order to understand structures which are otherwise too abstract [26]], or when we are in presence of *redundant* and unmanageable information; 2. action can be useful in presence of *incomplete* or *inconsistent* information – not only from the "perceptual" point of view – or of a diminished capacity to act upon the world: it is used to get more data to restore coherence and to improve deficient knowledge; 3. action enables us to build *external artifactual models* of task mechanisms instead of the corresponding internal ones, that are adequate to adapt the environment to the agent's needs. 4. action as a *control of sense data* illustrates how we can change the position of

"primitive" settings. Any written procedure is a simple example of a cognitive "mediating structure" with possible cognitive aims, so mathematical symbols and diagrams: "Language, cultural knowledge, mental models, arithmetic procedures, and rules of logic are all mediating structures too. So are traffic lights, supermarkets layouts, and the contexts we arrange for one another's behavior. Mediating structures can be embodied in artifacts, in ideas, in systems of social interactions [. . .]" [21, pp. 290–291].

9 It is difficult to preserve precise spatial and geometrical relationships using mental imagery, in many situations, especially when one set of them has to be moved relative to another.

our body (and/or of the external objects) and how to exploit various kinds of prostheses (Galileo's telescope, technological instruments and interfaces) to get various new kinds of stimulation: action provides some tactile and visual information (e.g., in surgery), otherwise unavailable.

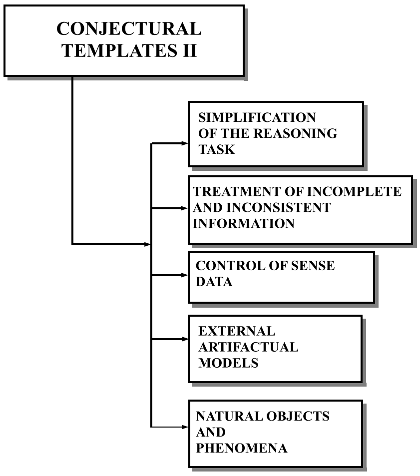

Figure 2. Conjectural templates II.

3.2 Mirroring Hidden Properties through Optical Diagrams

An interesting epistemological situation is the one concerning the cognitive role played by some special epistemic mediators in the field of non-standard analysis, an "alternative calculus" invented by Abraham Robinson [27], based on infinitesimal numbers in the spirit of Leibniz method. It is a kind of calculus that uses an extension of the real numbers system $\mathbb{R}$ to the system $\mathbb{R}^*$ containing infinitesimals smaller in the absolute value than any positive real number. I maintain that in mathematics diagrams play various roles in a typical abductive way. Two of them are central:

- they provide an intuitive and mathematical *explanation* capable of facilitating the understanding of concepts difficult to grasp, that appear hidden, obscure, and/or epistemologically unjustified, or that are *not expressible* from an intuitive point of view;
- they help *create* new previously unknown concepts.

In the construction of mathematical concepts many external representations are exploited, both in terms of diagrams and of symbols. I am interested in my research in diagrams which play an *optical* role – microscopes (that look at the infinitesimally small details)[28], telescopes (that look at infinity), windows (that look at a particular situation), a *mirror* role (to externalize rough mental models), and an *unveiling* role (to help create new and interesting mathematical concepts, theories, and structures).

The role of an "optical microscope" that shows the behavior of a tangent line is illuminating. In standard analysis, the change dy in y along the tangent line is only an approximation of the change Δy in y along the curve. But through an optical microscope, that shows infinitesimal details, we can see that $dy = \Delta y$ and then the quotient $\Delta y/\Delta x$ is the same of dy/dx when $dx = \Delta x$ is infinitesimal (see Figure 3 and, for details, [28]). This removes some difficulties of the representation of the tangent line as limit of secants, and introduces a more intuitive conceptualization: the tangent line "merges" with the curve in an infinitesimal neighborhood of the contact point.

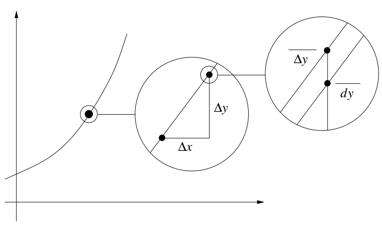

Figure 3. An optical diagram shows an infinitesimal neighborhood of the graph of a real function.

Only through a second more powerful optical microscope "within" the first (I call this kind of epistemic mediators *microscopes within microscopes*), we can see the difference between the tangent line and the curve. Under the first diagram, the curve looks like the graph of

$$f'(a)x,$$

i.e., a straight line with the same slope of its tangent line;[10] under the second, the curve looks like

$$f'(a)x - \frac{1}{2}f''(a).$$

This suggests nice new mental representations of the concept of tangent line: through the optical lens, the tangent line can be seen as the curve, but through a more powerful optical lens the graph of the function and the graph of the tangent are distinct, straight, and parallel lines. The fact that one line is either below or above the other, depends on the sign of $f''(a)$, in accordance with the standard real theory: if $f''(x)$ is positive (or negative) in a neighborhood, then f is convex (or concave) here and the tangent line is below (or above) the graph of the function.

However, this easily mirrors a sophisticated *hidden* property. Let f be a two times differentiable function and let a be a flex point of it. Then $f''(a) = 0$ and so the second microscope shows again the curve as the same straight line: this means that the curve is "very straight" in its flex point a. Of course, we already know this property – the

[10] This is mathematically justified in [28].

curvature in a flex point of a differentiable two times function is null – which comes from standard analysis, but through optical diagrams we can find it immediately and more easily (the standard concept of curvature is not immediate).

Some diagrams could also play an unveiling role, providing new light on mathematical structures: it can be hypothesized that these diagrams can lead to further interesting creative results.

Let us now consider an unveiling diagram in the Lobachevskyan discovery of the elementary non-Euclidean geometry (cf. Figure 4 [29]).

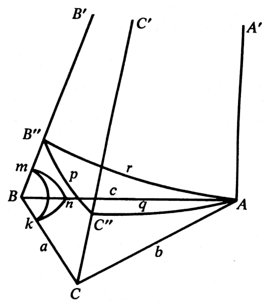

Figure 4. Unveiling diagram.

This diagram exploits "audacious" representations in the perspective of three dimensional geometrical shapes. The construction given in the figure aims at diagrammatically "representing" a stereometric non-Euclidean form built on a rectilinear right angled triangle ABC to which theorems previously proved (for example, the one stating that the parallels AA', BB', CC', which lie on the three planes are parallels in non-Euclidean sense) can be applied. In this way Lobachevsky is able to further apply symbolic identifications and to arrive to new equations which consistently (and in the same time) connect Euclidean and non-Euclidean perspectives. This kind of diagram strongly guides the geometer's selections of moves by eliciting what I call the Euclidean-inside non-Euclidean "model matching strategy". This maneuver also constitutes an important step in the affirmation of the modern "scientific" concept of model. This unveiling diagram constitutes a kind of gateway to imaginary entities.[11]

I stated that in mathematics diagrams play various roles in a typical abductive way. We can add that:

- they are *epistemic mediators* able to perform various abductive tasks
- they are *external representations* which provide explanatory and abductive results also fruitful in some aspects of chance production.

[11] More details concerning the role of mirror and unveiling diagrams in the discovery of non-Euclidean geometry in Lobachevsky's thought are given in [25].

4 Mechanizing Manipulative Abduction

4.1 Geometrical Construction is a Kind of Manipulative Abduction

Let's quote an interesting passage by Peirce about constructions. Peirce says that mathematical and geometrical reasoning "consists in constructing a diagram according to a general precept, in observing certain relations between parts of that diagram not explicitly required by the precept, showing that these relations will hold for all such diagrams, and in formulating this conclusion in general terms. All valid necessary reasoning is in fact thus diagrammatic" [16, 1.54]. Not dissimilarly Kant says that in geometrical construction "[. . .] I must not restrict my attention to what I am actually thinking in my concept of a triangle (this is nothing more than the mere definition); I must pass beyond it to properties which are not contained in this concept, but yet belong to it" [30, A718-B746, p. 580].

We have seen that manipulative abduction is a kind of usually model-based abduction, that exploits external models endowed with delegated (and often implicit) cognitive roles and attributes. 1. The model (diagram) is external and the strategy that organizes the manipulations is unknown a priori. 2. The result achieved is new (if we, for instance, refer to the constructions of the first creators of geometry), and adds properties not contained before in the concept (the Kantian to "pass beyond" or "advance beyond" the given concept [30, A154-B194, p. 192].[12]

Humans and other animals make a great use of perceptual reasoning and kinesthetic abilities. We can catch a thrown ball, cross a busy street, read a musical score, go through a passage by imaging if we can contort out bodies to the way required, evaluate shape by touch, recognize that an obscurely seen face belongs to a friend of ours, etc. Usually the "computations" required to achieve these tasks are not accessible to a conscious description. Mathematical reasoning uses language explanations, but also non-linguistic notational devices and models. Geometrical constructions represent an example of this kind of extra-linguistic machinery we know as characterized in a model-based and manipulative - abductive - way.

4.2 Automatic Geometrical Constructions as Extra-Theoretical Epistemic Mediators

An example of mechanization of manipulative abduction is ARCHIMEDES, a very interesting artificial intelligence computer program [31,32] that represents geometrical diagrams (points, line segments, polygons, and circles) both as pixels arrays and as propositional statements.[13] For example a triangle will be represented from a propositional description as a set of marked pixels in an array, together with a set of data naming the given triangle and storing facts about it (for instance that it is right) and

[12] Of course in the case we are using diagrams to demonstrate already known theorems (for instance in didactic settings), the strategy of manipulations is not necessary unknown and the result is not new.

[13] This approach in computer science, involving the use of diagram manipulations as forms of acceptable methods of reasoning, was opened by Gelernter's Geometry Machine [33], but the diagrams played a very secondary role.

constraints upon it (perhaps that it remains right throughout this use of the diagram). Hence, a computational equivalent of a physical diagram is represented, plus some human propositional knowledge about it.

The program is able to manipulate and modify its own representations of diagrams[14], that is it is able to make geometrical constructions (called "simulation constructions"): adding parts or elements, moving components about, translating and rotating by preserving metric properties, of course subordinated to the given specific constraints and to the whole structure of the two-dimensional space. Some knowledge of algebra is added, and of the taxonomic hierarchy of geometric figures (all squares are rectangles, etc.); moreover, it is also added additional knowledge like side-angle-side congruency theorem and the sum of the interior angles of a triangle, knowledge of problem solving strategies and heuristics, knowledge of logic (for example: a universal statement can be disproved by a single counterexample) [34].

When the program manipulates the specific diagram, it records the new information that comes out, then it can for example detect sets of area equivalences, and so on: for example, it is able to verify that a demonstration of the Pythagorean Theorem is correct, mirroring its truth in terms of constructions and manipulations. To account for the universality of geometrical theorems and propositions many different methods for learning and "generalizing" the specific instance of the constructed diagram are exploited ([32, pp. 260-264]). These methods come from a kind of predicative knowledge "exogenous" of course to the mere diagrammatic representation: generalization is not a possible product of the pure diagrammatic understanding.

4.3 Automatic "Thinking through Doing"

Geometric constructions are certainly epistemic mediators that exploit the semantics of two dimensional diagrams (rather than the syntax of formal propositions) to perform various manipulative abductive tasks (discover a new property or new proposition/hypothesis, selecting suitable sequences of constructions as able to convincingly verifying theorems, etc).

Geometrical construction, one of the most ancient exploitations of two-dimensional diagrams for both practical and mathematical problem solving, is in turn "embodied" in a computational program, that is, finally, in a machine.[15]. From the epistemological point of view it is important to note that the program shows how it is possible to delineate the rules and the procedures that underlie the diagrams as models of propositions about space, that is able to capture the structure of space. The kind of reasoning described is very rich and takes advantage of almost all the resources of two-dimensional space (going beyond the simple use of topological properties like in the case of Euler/Venn diagrams). Moreover, the physical diagram necessarily (that is for the objective reasons of its materiality) preserves topological and geometric properties of two-dimensional space.

[14] For details cf. [34]

[15] Humans can think using geometrical constructions also without "doing", for instance in the case of "thinking through drawing" at the level of imagination. Kant (and Proclus) were perfectly clear when they referred to the role of imagination as the condition of possibility of the empirical drawing itself [10].

We know that the structure of intuitive space is also embraced by analytic geometry. Lindsay observes that, in general, it could be better to use an analytic representation because conventional digital computers are "a natural match for numerical representation". If we consider the various ways of representing geometrical diagrams and their behavior - the analytic representation is an example - it is important to point out their real cognitive nature. I think we have to agree with the following position: "This does not mean, however, that diagrams represented numerically are not really diagrams. What makes them diagrams is not bits or voltages or axioms or CCD signals. What makes them diagrams is that they capture the structure of space. This is another way of saying that they enforce constraints on the behavior of the representations that reflect restrictions on the behavior of objects in space" (*ibid.*).

5 Conclusions

We have seen that, to solve the problem of the so-called "logic of discovery", we need to clarify what we are looking for, and the meaning of concepts like *creativity* and *discovery*. Following Peircian ideas, we see the recent computational modeling as very useful in a strictly pragmatical sense. We can produce and implement actual, and then possible, *rational* models of creative reasoning and scientific discovery. In this intellectual framework a new paradigm, aimed at unifying the different perspectives, is played by the fundamental concept of *abduction*. Many "working" abductive processes can be found and studied that are rational, unambiguous, and perfectly communicable. I have maintained that the concepts of *model-based* and *manipulative* abduction are important not only in delineating the actual practice of abduction, but also in the development of programs computationally adequate to rediscover, or discover for the first time, for example, scientific hypotheses and mathematical theorems or laws.

References

1. Lakatos, I.: *Proofs and Refutations. The Logic of Mathematical Discovery*. Cambridge University Press, Cambridge (1976)
2. Boden, M.A.: *The Creative Mind: Myths and Mechanisms*. Basic Books, New York, NY (1991)
3. Sternberg, R.J., Kaufman, J.C., Pretz, J.E., eds.: *The Creativity Conundrum : A Propulsion Model of Kinds of Creative Contributions*, New York, NY, Psychology Press (2002)
4. Buchanan, B.: Creativity at the metalevel. *AI Magazine* **77** (2001) 13–28
5. Kuhn, T.S.: *The Structures of Scientific Revolutions*. University of Chicago Press, Chicago (1962)
6. Reichenbach, H.: *Experience and Prediction*. University of Chicago Press, Chicago, IL (1938)
7. Popper, K.R.: *The Logic of Scientific Discovery*. Hutchinson, London, New York (1959)
8. Peirce, C.S.: *Writings of Charles Sanders Peirce, A Chronological Edition*, Vol. 3 (1872-1878). Indiana University Press, Bloomington, IN (1986)
9. Magnani, L.: *Abduction, Reason, and Science. Processes of Discovery and Explanation*. Kluwer Academic/Plenum Publishers, New York (2001)
10. Magnani, L.: *Philosophy and Geometry. Theoretical and Historical Issues*. Kluwer Academic Publisher, Dordrecht (2001)

11. Bylander, T., Allemang, D., Tanner, M.C., Josephson, J.R.: The computational complexity of abduction. *Artificial Intelligence* **49** (1991) 1991
12. Boutilier, C., Becher, V.: Abduction as belief revision. *Artificial Intelligence* **77** (1995) 43–94
13. Magnani, L.: Creative abduction and hypothesis withdrawal in science. In Meheus, J., Nickles, T., eds.: *Proceedings of the International Conference on Discovery and Creativity*, Dordrecht, Kluwer Academic Publishers (forthcoming)
14. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA (1998)
15. Flach, P., Kakas, A., eds.: *Abductive and Inductive Reasoning: Essays on Their Relation and Integration*, Dordrecht, Kluwer Academic Publishers (2000)
16. Peirce, C.S.: *Collected Papers*. Harvard University Press, Cambridge, MA (CP) 1–6, ed. by C. Hartshorne and P. Weiss, 7–8, ed. by A. W. Burks, 1931-1958.
17. Magnani, L.: Inconsistencies and creative abduction in science. In: *AI and Scientific Creativity. Proceedings of the AISB99 Symposium on Scientific Creativity*, Edinburgh, Society for the Study of Artificial Intelligence and Simulation of Behaviour, University of Edinburgh (1999) 1–8
18. Thagard, P.: *Conceptual Revolutions*. Princeton University Press, Princeton (1992)
19. Nersessian, N.J.: Should physicists preach what they practice? Constructive modeling in doing and learning physics. *Science and Education* **4** (1995) 203–226
20. Nersessian, N.J.: Model-based reasoning in conceptual change. In Nersessian, N.J., Magnani, L., Thagard, P., eds.: *Model-based Reasoning in Scientific Discovery*, New York, Kluwer Academic/Plenum Publishers (1999) 5–22
21. Hutchins, E.: Cognition in the Wild. MIT Press, Cambridge, MA (1995)
22. Nersessian, N.J.: How do scientists think? Capturing the dynamics of conceptual change in science. In Giere, R., ed.: Cognitive Models of Science. Minnesota Studies in the Philosophy of Science, Minneapolis, University of Minnesota Press (1992) 3–44
23. Gooding, D.: *Experiment and the Making of Meaning*. Kluwer, Dordrecht (1990)
24. Norman, D.A.: *Things that Make Us Smart. Defending Human Attributes in the Age of the Machine*. Addison-Wesley, Reading, MA (1993)
25. Magnani, L.: Epistemic mediators and model-based discovery in science. In Magnani, L., Nersessian, N.J., eds.: *Model-Based Reasoning: Science, Technology, Values*, New York, Kluwer Academic/Plenum Publishers (2002) 305–329
26. Piaget, J.: *Adaption and Intelligence*. University of Chicago Press, Chicago (1974)
27. Robinson, A.: *Non-Standard Analysis*. North Holland, Amsterdam (1966)
28. Magnani, L., Dossena, R.: Perceiving the infinite and the infinitesimal world: unveiling and optical diagrams and the construction of mathematical concepts (2004) Forthcoming in *Foundations of Science*.
29. Lobachevsky, N.J.: *Geometrical Researches on the Theory of Parallels*. University of Texas, Austin, TX (1891)
30. Kant, I.: *Critique of Pure Reason*. MacMillan, London (1929) translated by N. Kemp Smith, originally published 1787, reprint 1998.
31. Lindsay, R.: Understanding diagrammatic demonstrations. In Ram, A., Eiselt, K., eds.: *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, Paris, Erlbaum, Hillsdale (1994) 572–576
32. Lindsay, R.: Using diagrams to understand geometry. *Computational Intelligence* **9(4)** (1998) 343–345
33. Gelertner, H.: Realization of a geometry theorem proving machine. In: *International Conference on Information Processing*, Paris, Unesco House (1959) 273–282
34. Lindsay, R.: Playing with diagrams. In M. Anderson, P.C., Haarslev, V., eds.: *Diagrams 2000*, Berlin, Springer (2000) 300–313

An Approach to Generating New Ideas Based on Linking the Frames of Concepts

Kambiz Badie and Maryam Tayefeh Mahmoudi

Info Society Dept., Iran Telecom Research Center
Tehran, Iran
Tel: +98 21 8497705
{k_badie, mahmodi}@itrc.ac.ir

Abstract. Being inspired by the methodology of conceptual integration, in this paper we propose an approach to generating new ideas, based on linking the frames of concepts. Regarding this, concepts are represented in terms of frames with appropriate attributes, and the values of these attributes are then linked with each other via applying some conjunction structures. In the paper, the application of the proposed approach is discussed for scientific frameworks as new ideas.

1 Introduction

Generating new ideas seems in many cases to depend on the status of the background knowledge developed in human mind within the course of his/ser personal experiences with the surrounding environment. It is expected that, a dominant part of this background knowledge be devoted to some previously examined concepts, which have been shown to be useful for a range of problems. In this sense, conceptual integration (or conceptual blending) seems to be highly workable for justifying different sorts of idea generation. Being inspired by this methodology, in this paper we propose an approach to generating new ideas, based on linking the frames of concepts. In this respect, concepts are represented in terms of the frames with appropriate attributes, and the values of these attributes are then linked with each other via applying some conjunction structures. Conjunction structures in our approach are defined to be compound in their nature. In the paper, the application of the proposed approach is discussed for generating the scientific frameworks as the alternative for new ideas.

2 Some Existing Approaches for Idea Generation

Different approaches for idea generation are classified into three major categories, from our viewpoint, as follows:

- Conceptual Integration (or Conceptual Blending)

Within this framework, traditional spaces associated with analogical or metaphoric mapping, the source and the target, combine via some structural mapping to produce another independent blended space that provides the local point for the resultant integration [1,2]. The project called “Dr. Divago” within which the use of a multi-domain knowledge base as a problem-solving resource is proposed, is the first trial for computationalization of conceptual integration [3,4]. When searching for a solution, Dr. Divago may diverge to a domain different form that in focus, through the use of a mapping function that works as a cross-domain bridge, or to a domain blend, i.e., a domain that results from blending the domain in focus with a different domain.

- Approaches Based on Analogical Reasoning/Case Based Reasoning

According to this category, final ideas, at different abstraction levels, are obtained through a sort of analogical mapping from the source onto the target, to describe why, how and when a pattern can be applied to a certain problem situation. Some of the approaches have been focused on designing software patterns through development of case tools, which are intended to aid the software engineer in the design phase, to retrieve the relevant designs by using unified mark up language to represent the design knowledge [5,6]. Also, regarding the fact that the analogical mapping may, in certain situations, need to be performed indirectly, according to some other approaches in this category, the concepts in the target are first mapped onto some generic concepts in an intermediate space, and the results are then reinterpreted into appropriate concepts in the target. To step toward potential interpretation schemes, the concepts in the intermediate space should be selected as generic as possible, and should be able to show the major mental modes the reasoner may exhibit within the reasoning process. [7,8].

- Integrated Model for Analogical Modeling

Within this model, creativity is interpreted as the search for some source analogue with which to re-interpret a given target domain. Individual stage in it eliminates the influence of interactions between stages. Providing the inspiration or a further retrieval episode, will deliver an all-encompassing explanation about this model [9].

3 Basics

The basic motive behind the proposed approach to linking the frames of concepts is to structure a framework for understanding how a structured concept A; a concept represented in a structured and hierarchical manner, can be applied to another structured concept B to (i) confront with the concept B's probable drawbacks / deficiencies, (ii) justify its own role /utility, or (iii) anticipate its own prospect with respect to the concept B. For instance, when in the domain of computation, "Fuzzy" as the concept A is applied to "Clustering" as the concept B, one expects that some drawbacks/deficiencies of Clustering may be removed through applying Fuzzy to it, thus leading to formation of Fuzzy Clustering as a composite concept with enhanced capabilities for clustering in certain situations. Also, when "Image Processing" as the Concept A is applied to "Medical Diagnosis" as the concept B, one expects to figure out finally the possible roles / utilities of Image Processing regarding Medical Diagnosis. In the meantime, prospect of deploying "Pollution Making Industries" in "Societies with Certain Ecological Conditions" is an issue, which can be anticipated by means of linking the two concepts in some manner. Linking the frames of concepts is thus expected to help innovation come about successfully in many situations.

Suppose that the concepts under study are some scientific concepts coming from the domains such as basic methodologies. As the concepts have been decided to be in terms of frames, it is first necessary to categorize them from the viewpoint of the similarities between the types of attributes. In our study, we have considered four groups of "action-type noun", "non-action type noun", "adjective" and "adverb" with respect to characterization of attributes.

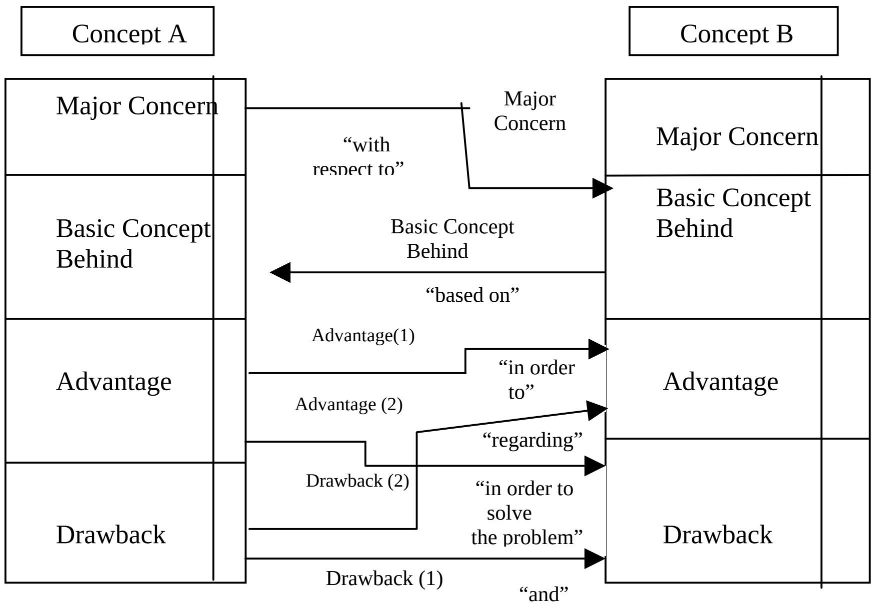

Fig. 1. Conjunction structures essential to linking the frames of action-type concepts

Table 1 shows the prime attributes we have developed for the two categories of "action-type noun" and "non-action type noun". To link the frames of concepts, it is necessary to link the values of the related attributes, using some appropriate conjunction structures preferably with compound nature. Role of a conjunction structure is to make a fluent and comprehensible conjunction between the related values. They might be varying based on the situation in these values, and it is thus important to consider the possibility of setting the conjunction structures on-line. The idea on the type of the conjunction structure and the formats for expressing it, depends on the role of each attribute. For instance, the value of the attribute Major Concern in the composite concept, should be determined through linking the value of the same attribute in the Concept A, with the value of the attribute "Basic Concept Behind" in the Concept B via using the conjunction structure "with respect to", or under certain circumstances "when it comes to". The conjunction structures, which are essential to linking the frames of action-type concepts, are illustrated in Figure1.

Table 1. The prime attributes for the "action-type noun" and "non-action type noun" categories

Action – type noun	Super Class, Sub Class, Instance, Major Concern. Basic Concept Behind, Related Non-Action Type Noun Entity, Related Action Type Noun Entity, Advantage, Disadvantage
Non – action type noun	Super Class, Sub Class, Instance, Major Role, Basic Component, Specification, Related Non-Action Type Noun Entity, Related Action Type Noun Entity, Pragmatic Scenario

4 Some Examples

4-1 Example 1: Linking the frames of "Fuzzy" and "Clustering"
An example is presented in Figure 2, where, two frames of "Fuzzy" (Figure 2(a)) and "Clustering" (Figure 2(b)) including the attributes discussed in 3, are linked in two different directions. (Figure2(c)) and Figure 2(d))

Each direction indicates the way the first concept is used to remove the drawbacks of the second concept. As it is seen from Figure 2(c) and Figure 2(d), "action-type nouns", "non-action type nouns", and the conjunction structures are used in appropriate forms to determine the final content of the corresponding attributes in the final composite concept.

For instance, the value of the attribute "Major Concern" in "Fuzzy Clustering" is determined through linking "Fuzzy" 's "Major Concern" with some prime items in the value of "Clustering" 's "Basic Concept Behind" via using the conjunction structure "with respect to", while the value of "Basic Concept Behind" for "Fuzzy Clustering" is determined through linking "Fuzzy" 's "Basic Concept Behind" with the value of "Clustering" 's "Basic Concept Behind" using "Based on" as the conjunction structure.

The reason for picking out the prime items in the value of the concept B's “Basic Concept Behind” (in case of determining the value of the attribute “Major Concern” in the composite concept), goes back to the related conjunction structure, which is “with respect to” in this case. When a phrase comes after such a conjunction structure, it is obvious that some prime items in this phrase are to be highlighted as the alternatives, with respect to which the entire sentence makes sense. In the example above, “sufficiently close points in the feature space” are the prime items in the value of the attribute “Basic Concept Behind” in the frame of Clustering as the Concept B. Let say, in this way, this sense will appear that, the major concern of the composite (or blended) concept of “Fuzzy Clustering”, would be to manage uncertainty with respect to those points, which are close enough to each other in the feature space. This together with the value of the attribute “Basic Concept Behind” in “Fuzzy Clustering” will eventually lead to the idea that Membership Functions (MFs) can be used to handle uncertainty regarding the distance between those points in the feature space, which are sufficiently close to each other. Also, regarding the composite (or blended) concept “Clustering Fuzzy”, the eventual idea is that, clustering techniques can be applied to MFs in order to turn them into more efficient structures for uncertainty handling purposes. In this case, a variety of ideas may subsequently be generated for clustering MFs, each depending on its own interpretation of MF as a pattern. One way is to regard MF as a geometric shape consisting of some geometrically significant points. Regarding this case, these points would become subject to clustering. The other way is to regard MF as a mathematical function like a Gaussian Distribution Function with some basic parameters such as variance and mean value. In this case, the total vector comprising of the values of these parameters would be a subject for clustering. Regarding the above discussion, application of Clustering as a technique, to Fuzzy (Processing) as another technique, may yield some new ideas as follow:

- Generating class structures for classifying the geometrically significant points of a fuzzy membership function.
- Generating class structures for classifying the (mean value, variance) vector, which represents the Gaussian Distribution for fuzzy membership function

This can be quite interesting from research & development viewpoint, since the researcher/developer can in this way become aware of the possibilities behind different compositions of the existing techniques, and choose the one that can best fit his/her current problem requirements.

Fuzzy **(a)**

Super Class	Logic of uncertainty
Sub Class	T-Norm,…
Major Concern	Uncertainty management
Basic Concept Behind	Utilizing MF to determine the status of assigning a pattern to different classes, Using some operators for decision making
Advantage	Enhancing DM performance through considering the affiliation of a pattern toward different classes
Drawbacks	MFs may not be defined properly

Clustering (b)

Super Class	Category theory
Sub Class	Model-based, adaptive, neural, self-organizing,…
Major Concern	To generate class structures for classification purposes
Basic Concept Behind	Determination of sufficiently close points in the feature space
Advantage	Optimizing search with respect to Decision Making
Drawbacks	Defining distinct classes

Fuzzy Clustering (c)

Super Class	Category theory
Sub Class	Fuzzy model-based clustering
Major Concern	Uncertainty management with respect to sufficiently close points in the feature space
Basic Concept Behind	Determination of sufficiently close points in the feature space based on utilizing MF to determine the status of assigning a pattern to different classes
Advantage	Enhancing Decision Making performance through considering the affiliation of a pattern toward different classes in order to optimize search, Enhancing … in order to solve the problem of distinct classes
Drawbacks	MFs may not be defined properly regarding Decision Making, MFs may not be defined properly and define distinct classes

Clustering Fuzzy (d)

Super Class	Logic of uncertainty
Sub Class	Clustering T-Norm
Major Concern	To generate class structures for classification with respect to MF
Basic Concept Behind	Utilizing MF to determine the status of assigning a pattern to different classes based on determination of sufficiently close points in the feature space
Advantage	Optimizing search with respect to Decision Making in order to solve the problem of MFs, Optimizing search with respect to Decision Making in order to Enhance DM
Drawbacks	Defining distinct classes may be problematic regarding the decision making process, Defining distinct classes may be problematic and MFs may not be defined properly

Fig. 2. Linking frames of the concepts "Fuzzy" and "Clustering"

It is interesting to notice that from conceptual integration or conceptual blending viewpoint, each frame participating in the linking, corresponds to a separate input space, while the attributes of the two frames constitute the generic space, and finally the conjunctions structures are propounded as the operators for blending the values coming from the input spaces, according to this generic space. The schematic for such a viewpoint is illustrated in Figure 3.

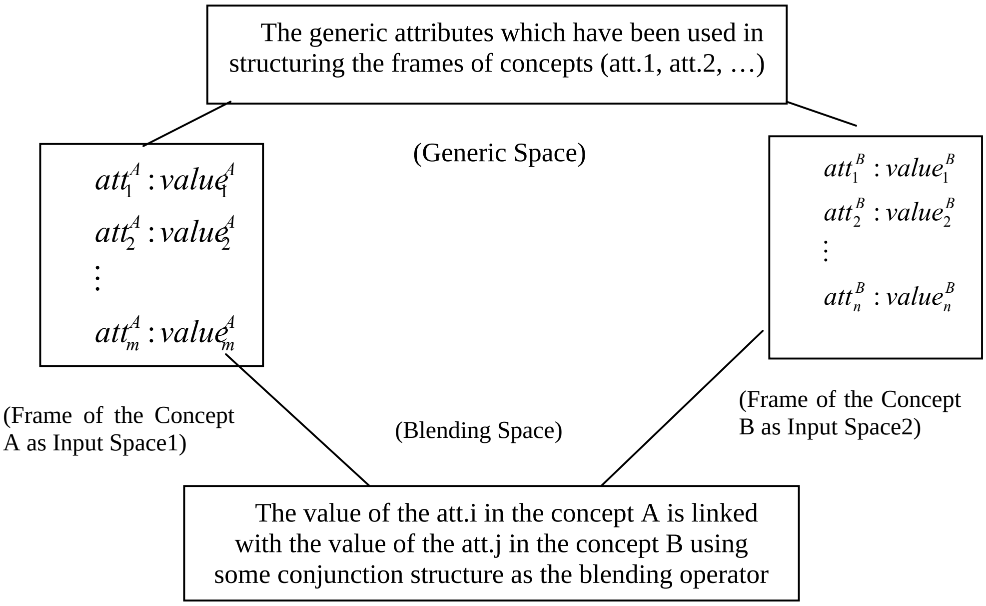

Fig 3. A schematic for representing the concept linking approach within the framework of conceptual integration

4-2 Example2: Linking the frames of theories for generating further ideas

In this example, the frame of a theory as a concept is linked with the frame of another theory, in order to identify further capabilities for the later in specific situations. Regarding this, theories frames should be equipped with appropriate attributes that can stand for their major aspects. In our approach, the attributes already discussed in 3, are equally used for the concepts of theories.

The area for our example is Cultural Anthropology. In this respect, we try to show how "Functionalism" as a major theory in Cultural Anthropology can help "Historical Particularism" as another major theory come up with further capabilities in specific situations. Taking the definitions of these two theories into consideration, the values for the attributes "Major Concern" and "Basic Concept Behind" are illustrated in Figure 4(a) and 4(b), respectively.

It is illustrated in the figure how linking the values of certain attributes in the two concepts through using the conjunction structures illustrated in Figure 1, can lead to determination of the appropriate values for the attributes in the composite concept.

As is seen from the content of the composite concept of Fig. 4, the conjunction structure “with respect to” in the value of the attribute “Major Concern”, has been replaced by “when it comes to”. The reason for such a decision, is the particular emphasis, which is to be made on the terms “incorporating”, “collection”, and “documenting”, which the values of the attribute “Basic concept Behind” in the frame of “Historical Particularism” start. Let say, these terms, which are action-type nouns, have enough potential (from their process viewpoint) to deserve coming after a conjunction structure like “when it comes to”, and not “with respect to”. It is interesting to notice that, through regarding the prime terms in the newly- developed contents as separate frames (as it was also discussed in 4-1), some contents may be obtained which are transparent and comprehensive enough to the researcher of this field. For instance, assuming that prehistory, linguistics and physical anthropology are the prime terms respecting the concept "Historical Particularism", through considering them as separate frame, the researcher may find out ultimately that the above value can be interpreted in the following manner: seeing society as a system of interrelated parts that operate independently, when it comes to incorporating preliminary style of living (obtained from expanding the frame of prehistory), patterns of interpersonal communication (obtained from expanding the frame of linguistics), and style of nutrition (obtained from expanding the frame of physical anthropology). The above interpretation may lead the researcher in the field to the point that a major concern behind applying Functionalism to Historical Particularism, can lie in the way that all the elements of prehistory, linguistics and physical anthropology in terms of "preliminary style of living", "interpersonal communication" and "style of nutrition", are considered in the process of seeing a society as a system of interrelated parts.

Regarding the above discussion, an interesting point, with respect to linking the frames of theories, is drawing the attention of the researcher to some specific viewpoints whose interpretation may finally lead to some new strategies, tactics and techniques for handling more complex cultural anthropological issues. For instance, results of applying the ideas of Functionalism to Historical Particularism in terms of the attributes' values, will offer some clues to the researchers, which, based upon interpretation, will yield new strategies and tactics that can help satisfy a broader range of expectations within Cultural Anthropology. In this way, two points are remarkable as follows:

- Idea generation can be regarded as a process of interpreting the content obtained through linking the frames of concepts
- To generate comprehensible ideas, this interpretation should be hierarchical, where at each level of hierarchy some new prime terms are regarded as new frames providing the possibility for generating new contents.

Functionalism **(a)**

…	…
Major Concern	• Seeing society as a system of interrelated parts that operate independently
Basic Concept Behind	• Particular social forms functioning from day to day, to reproduce the traditional structure of the society
…	…

Historical Particularism (b)

…	…
Major Concern	• Emphasizing on particulars as opposed to universals • Focusing on individuals • Challenging ethnocentric assumptions of the environments
Basic Concept Behind	• Incorporating prehistory, linguistics, and physical anthropology • Careful collection of ethnographical data and field experience (as opposed to the comparative methods of unilineal cultural evolutionists) • Documenting and studying vanishing cultures
…	…

Functional Historical Particularism (c)

…	…
Major Concern	• Seeing society as a system of interrelated parts that operate independently), <u>when it comes to</u> incorporating prehistory, linguistics, and physical anthropology • Seeing society as a system of interrelated part (that operate independently), <u>when it comes to</u> collection of ethnographical data and field experience (as opposed to the comparative methods of unilined cultural evolutionists) • Seeing society as a system of interrelated parts (that operate independently), <u>when it comes to</u> documenting and studying vanishing cultures
Basic Concept Behind	• Incorporating prehistory, linguistics, and physical anthropology, <u>based on</u> particular social forms functioning from day to day to reproduce the traditional structure of the society • Careful collection of the anthropological data and field experience, <u>based on</u> particular social forms functioning from day to day to reproduce the traditional structure of the society • Documenting & studying vanishing cultures, <u>based on</u> particular social forms functioning from day to day to reproduce the traditional structure of the society
…	…

Fig. 4. Linking frames of theories for generating further ideas

5 Assessing the Concept Linking Approach to Idea Generation from the Viewpoint of Conceptual Integration

As it was discussed, each frame of concept participating in the linking corresponds to a separate input space, while the generic attributes which are to be used in representing the contents, constitute a generic space, according to which blending operators in terms of some conjunction structures link the corresponding values, to make the content for the final composite concept.

The entire process is fully purpose-directed, and no divergent thinking or randomness, such as the one observed in case of Dr Divago project, exists respecting such a process of linking. In other words, the effect of diversity in the final solution, in the way Dr Divago project expects for, is not expected here.

It should however be mentioned that, due to a variety of reasons, some sort of variation may take place in the process of linking, which is briefly discussed:

1. The prime items in the value of an attribute (as discussed in 4-1) may be multiple.
2. Multiplicity in the values of an attribute. For instance, according to Fig. 2(a), the attribute "Basic Concept Behind" has two values, which can both participate in concept linking.
3. The entire process of linking may be performed hierarchically in the sense that, once the values of two attributes (one in the concept A and the other in the Concept B) were linked via using an appropriate conjunction structure, possibility would be left for the prime items in the resulted values to propound themselves as some new frames whose values can in some way appear at the next stages of the linking process.

For instance, referring to the attribute "Basic Concept Behind" of the composite concept "Fuzzy Clustering", whose value, as the result of the first stage in the idea generation process, has been decided to be "Determination of sufficiently close points in the feature space based on utilizing MF", the term MF (which is a non-action type noun) itself can be propounded as a frame, which according to Table 1, can hold the attributes such as "instance", "specification", etc. Taking this point into account, the term MF in the above-mentioned phrase can be substituted by the terms such as "Gaussian Distribution Function" (as a choice for the value of the attribute "Instance"), "Linguistic Variable Value" (as a choice for the value of the attribute "Related Non-Action Type Entity", and so on.

In this manner one may expect that a variety of explanations may be generated as the result of linking the contents of the two concepts, each emphasizing on a particular aspect.

Although these aspects are semantically overlapped, the entire process does not violate the idea generation process, mostly due to the fact that, similar novelties may be innovated in different ways and formats, based on the conditions and the situations in their ideators.

6 Concluding Remarks

It was discussed in the paper that linking the frames of concepts via using appropriate conjunction structures can lead to some new contents that can guide researchers to some specific viewpoints regarding new strategies, tactics and techniques for handling more complex situations in their domain. Within this scope, viewing the prime items in the generated content as new frames and replacing these items by the values obtained from these frames, can lead to generation of more detailed ideas. This can provide a suitable environment for the users to become familiar with a variety of ideas, as different interpretations of the way the frame of the concept can be linked with another.

Conclusion is that, linking the frames of concepts can achieve a promising role in interpreting the way concepts can mutually help each other, to finally lead to new methodologies and frameworks. This can have a lot of applications in a wide range of issues such as education/pedagogy, research & development, and managerial planning as well.

References

1. Fauconnier, G.,Turner, M.: Conceptual integration networks. Cognition Science, Vol.22, No. 2, (1998) 133-87
2. Fauconnier, G.,Turner, M.: Conceptual integration networks, in Jean-Piere Koening (Ed.), Discourse and Cognition, Center for the study of language and Information (CSLI) stanford, (1998) 155-70
3. Pereira, F.C., Cardoso, A.: Conceptual blending and the quest for the holy creative process, Proc. of the 2nd Workshop on Creative Systems: Approaches to Creativity and Cognitive Science, ECAI 2002, France, (2002)
4. Pereira, F.C., Cordoso, A.: The boat-house visual blending experiment, Proc. of the 2nd Workshop on Creative Systems: Approaches to Creativity and Cognitive Science, ECAI 2002, France, (2002)
5. Gomes, G., Pereira, F.C., Paiva, P., Seco, N., Carreiro, P., Ferreira, J.L., Bento, C.: Experiments on Case-Based Retrieval of Software Design, 6th European Conf. ECCBR2002, UK, Springer, (2002) 118-132
6. Gomes, P., Pereira, F.C., Paiva, P., Seco, N., Carreiro, P., Ferreia, J.L., Bento, C.: Using CBR for automation of Software Design patterns, 6th European Conf., ECCBR2002, Springer, UK, (2002) 534-548
7. Badie, K., Hejazi, M.: Creative idea generation via passing through an intermediate space between the source and the target, Proc. Of the workshop program at the forth Intl. Conf. On Case-Based Reasoning Canada, (2001) 125-128
8. Badie, K.: Creative idea generation via interpretative approach to analogical reasoning, Kybernetes, the Intl. Journal of System & Cybernetics, Vol. 31, No. 9/10, (2002)
9. O'Donoghue, D.: Towards a computational model of creative reasoning, Proc. Of Conf. On Computational Models of Creative Cognition, Ireland, (1997)