Proceedings of the Workshop
“Computational Creativity, Concept Invention,
and General Intelligence”

PICS
Publications of the Institute of Cognitive Science
Volume 2-2013
Computational Creativity, Concept Invention, and General Intelligence

2nd International Workshop, C3GI @ IJCAI 2013, Beijing, China, August 04, 2013

Proceedings
Volume Editors

Tarek R. Besold
Institute of Cognitive Science
University of Osnabrueck

Kai-Uwe Kühnberger
Institute of Cognitive Science
University of Osnabrueck

Marco Schorlemmer
IIIA-CSIC, Barcelona

Alan Smaill
School of Informatics
University of Edinburgh

This volume contains the proceedings of the workshop “Computational Creativity, Concept Invention, and General Intelligence (C3GI)” at IJCAI-2013.
Preface

This volume contains the proceedings of the 2nd International Workshop for Computational Creativity, Concept Invention, and General Intelligence held in conjunction with IJCAI 2013 in Beijing, China.

Continuing the mission started by the 1st edition of the workshop series in 2012 at Montpellier, the aim of this workshop is to bring together researchers from different fields of AI working on computational models for creativity, concept formation, concept discovery, idea generation, and their overall relation and role to general intelligence as well as researchers focusing on application areas, like computer-aided innovation.

Although the different approaches to questions concerning the aforementioned aspects do share significant overlap in underlying ideas, the cooperation between the respective communities is still in an early stage, and can greatly profit from interaction and discussion between people from the respective fields, forming trans- and interdisciplinary alliances in research and application.

Together with a growing community of researchers in both, academia and industry, we are convinced that it is time to revitalize the old AI dream of “Thinking Machines” which now has been almost completely abandoned for decades. Fortunately, more and more researchers presently recognize the necessity and feasibility of returning to the original goal of creating systems with human-like intelligence. By continuing the discussions started one year ago at the 2012 edition of the workshop, we are now working towards establishing the workshop series as an annual event gathering leading researchers in computational creativity, general intelligence, and concept invention whose work is in direct connection to our overall goal, providing and generating additional valuable support in further turning the current spirit of optimism into a lasting research endeavor.

August, 2013

Tarek R. Besold
Kai-Uwe Kühnberger
Marco Schorlemmer
Alan Smaill
Program Committee

Committee Co-Chairs

- Tarek R. Besold, University of Osnabrueck
- Kai-Uwe Kühnberger, University of Osnabrueck
- Marco Schorlemmer, IIIA-CSIC, Barcelona
- Alan Smaill, University of Edinburgh

Committee Members

- John Barnden, University of Birmingham
- Mehul Bhatt, University of Bremen
- Mark Bishop, Goldsmiths, University of London
- Simon Colton, Goldsmiths, University of London
- Gaetano Cascini, Politecnico di Milano
- Joerg Cassens, University of Luebeck
- Hamid Ekbia, Indiana University Bloomington
- Markus Guhe, University of Edinburgh
- Helmar Gust, University of Osnabrueck
- Ivan M. Havel, Charles University Prague
- Bipin Indurkhya, IIIT Hyderabad/AGH University of Science and Technology, Krakow
- Anders Kofod-Petersen, NTNU Trondheim
- Ulf Krumnack, University of Osnabrueck
- Oliver Kutz, University of Bremen
- Maricarmen Martínez Baldares, University of the Andes, Bogota
- Alison Pease, University of Edinburgh
- Francisco Pereira, University of Coimbra
- Enric Plaza, IIIA-CSIC, Barcelona
- Thomas Roth-Berghofer, University of West London
- Ute Schmid, University of Bamberg
- Petros Stefaneas, National Technical University Athens
- Tony Veale, University College Dublin
- Pei Wang, Temple University Philadelphia
Table of Contents

Full presentations:

Creativity in Artificial Intelligence as a Hybrid of Logic and Spontaneity  
S. Ellis, N. Sundar G., J. Valerio, S. Bringsjord & J. Braasch

How Models of Creativity and Analogy Need to Answer the Tailorability Concern  
J. Licato, S. Bringsjord & N. Sundar G.

Creative Concept Invention and Representation in an Interactive Child-Robot Healthy Snack Design Game  
A. Williams, A. Stroud, J. Panka & J. Williams

Compact presentations:

Back-and-Forth Inception: Towards a Cognitively-Inspired Model of Concept Construction via Conceptual Intellection  
A. Abdel-Fattah & S. Schneider

Square, zero, kitchen: Start – Insights on cross-linguistic conceptual encoding  
F. Quattri

Semantic Clues for Novel Metaphor Generator  
R. Rzepka, P. Dybala, K. Sayama & K. Araki
Creativity in Artificial Intelligence as a Hybrid of Logic and Spontaneity

Simon Ellis ● Naveen Sundar G. ● Joseph Valerio ● Selmer Bringsjord ● Jonas Braasch
elliss5@rpi.edu ● govinn@rpi.edu ● valerj3@rpi.edu ● selmer@rpi.edu ● braasj@rpi.edu
Department of Cognitive Science ● Department of Computer Science
The Rensselaer AI & Reasoning (RAIR) Lab ● Rensselaer Polytechnic Institute (RPI) ● Troy NY 12180 USA

Abstract

This paper explores machine creativity from the perspective that genuine human-level creativity is a hybrid function of formal, computational logic on the one hand, and intuition and spontaneity on the other. After setting the context and explaining the resources upon which we draw, we provide an example of this essentially synergistic relationship in the form of the improvisational functionality of the CAIRA system.

1 Introduction

After setting the context, this paper reports on new developments to our intelligent music agent, Handle, which was first introduced at C3GI 2012 in Montpellier, France. Handle is intended to operate as an artificial conductor, and to that end was developed to interpret and provide feedback on a human performance of a score-based, classical solo piece of music. The music calculus, a first-order sorted modal logic, was introduced as a means of evaluating certain aspects of musical performance, and cognitive attitudes directed thereto. The current version of Handle extends the work done in the previous version and explores similar questions in the area of improvised “free music.” We demonstrate Handle together with its sister agent FILTER as part of a larger musical system, CAIRA, and explore the intersections of logic, intuition, and spontaneity that together enable the creative process.

The plan of the paper is as follows. We begin by setting the context (§1). We note that in the present paper our AI work on computational creativity reflects a Bringsjordian perspective (§2.1), explain two assumptions that serve to place logic at the heart of our engineering (§2.2); give an overview of the “music calculus,” the logic-based framework used for our modeling (§2.3); and adumbrate our general approach to the nexus of logic and intuition (§2.4). Next (§3), we give a barbaric summary of our CAIRA and its two main components, and then, in section §4, we provide a demonstration, one that is broad but, on the logic side, thin (we end this section with a look at a deeper use of the music calculus). The paper wraps up with a brief discussion of next steps (§5).

2 Setting the Context

2.1 Psychometric View of Creativity Presupposed

In earlier work, we discussed several ways an artificial agent may via performance on certain tests at least seem creative, even though it may not be genuinely so. This test-based orientation reflects what is known as Psychometric AI [Bringsjord and Schimanski, 2003, Bringsjord, 2011]. The toughest relevant test is the “Lovelace Test,” which is passed by a computing machine when the machine’s output cannot be anticipated by its designer [Bringsjord et al., 2001]; a machine able to pass this test is said to be creativeB. We concede that this standard exceeds what our current engineering can reach. We have a more humble aim at present: engineer a computing machine able to convince non-experts, Turing-test style, that the output they observe is produced by what they would call a “creative” computing machine; this is creativeF. In the literature, there are even more latitudinarian tests for creativity: Cope, for instance, in his Computer Models of Musical Creativity [2005] tells us that for him creativity is “[t]he initialization of connections between two or more multifaceted things, ideas, or phenomena hitherto not otherwise considered actively connected” [Cope, 2005], where the evidence for such connections being made is provided by performance on certain simple puzzles.

2.2 Two Assumptions that Imply Use of Logic

Despite the fact that we anchor our r&d to tests, we do make one assumption about the internals of a creativeF machine, and one general assumption about the epistemic context of any computational artifact that we produce.

Regarding internals, we assume that the computing machine that is a candidate for creativityF have at least the minimum representation-and-reasoning power of quantified epistemic logic. This minimality condition, and the methodology that accompanies it, have been defended elsewhere [Bringsjord, 2008]. The condition reflects Bringsjord’s affirmation of Piaget’s seminal position that mature, general-purpose human cognition (including, most prominently, problem-solving) consists of processes operating largely on formulas having at least the expressivity of formulæ in full first-order logic [Inhelder and Piaget, 1958]. Given the affirmation of the condition in question, the present work reflects a desire to engineer machines that are creativeF: they
are both creative and their internal processing conforms to Piagetian concepts of general intelligence and creative problem-solving. For more on this machines that embody the first condition, see [Bringsjord and Licato, 2012].

What is our second assumption? We assume that computational creativity cannot be formalized and engineered without yet another nod in the direction of logic, and specifically epistemic logic. To see our second assumption, consider a person/agent A who has produced an artifact a through some creative process p. Any formal model of A’s cognition before, during, and after the creative process should by our lights have sufficient representational capacity to let us conclude, or at least consider, whether:

1. A knows that A itself has produced a in a creative manner.
2. A believes that a has not been produced by any other agents.
3. A knew what existed and what did not exist before it started the process that produced a.
4. A desired that it needed to create an artifact to fulfill some need.
5. A intended to engage in a series of actions making up the process p.

The above (informal) micro-theory to us is admittedly incomplete, but does seem unavoidable, and suggests that higher cognition surrounding creativity that can be succinctly captured with an expressive formal logic. We do not claim here that we have invented and specified this logic, but we have taken appreciable steps toward doing so, by way of our modeling framework, DCEC*, to which we now turn.

2.3 Our Logic-Based Framework (music calculus)

Our core logical framework is DCEC* (deontic cognitive event calculus) a multi-sorted quantified modal logic that has a well-defined syntax and a proof calculus. The syntax of the language of DCEC* and the rules of inference for its proof calculus are shown in Figure 1. DCEC* syntax includes a system of sorts S, a signature f, a grammar for terms t, and a grammar for sentences φ; these are shown on the left half of the figure. The proof calculus is based on natural deduction [Jaśkowski, 1934], and includes all the standard introduction and elimination rules for first-order logic, as well as additional rules for the modal operators; the rules are listed in the right half of the figure.

The music calculus is a specialization of DCEC* with logical symbols for reasoning over music, and was introduced in earlier work [Ellis et al., 2012], where it is discussed in some detail. Figure 2 shows a part of the signature of the music calculus used in the present paper. (We omit parts of the calculus not necessary for this paper.) The logic-based part of Handle is based on the music calculus.

More generally, the logical framework of DCEC* lets us model and compare different theories and implementations of creativity, intuition, and cognition in a precise manner. As explained below, the model of intuition used in this paper is based on the cost of computation. The agent needs to compute a particular value S(t), the state of the performance at the current time. The agent knows that there is a certain action αdeduce which when performed will lead to the agent simply believing that the value is something. There are costs associated with both the actions, and the agent would prefer knowing the value. More generally, if the agent believes that the cost of a full non-intuitive process is greater than the cost of an intuitive process by a threshold γ (which factors in the benefits of a deductive search), the agent would engage in an intuive process. Assuming that the agent knows some arithmetic, we can represent this model of intuition with this micro-theory:

Figure 1: Syntax of DCEC*

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Object</th>
<th>Agent</th>
<th>Sell ⊆ Agent</th>
<th>ActionType</th>
<th>Event</th>
<th>Moment</th>
<th>Boolean</th>
<th>Fluent</th>
<th>Numeric</th>
</tr>
</thead>
<tbody>
<tr>
<td>S ::=</td>
<td>Agent</td>
<td>×</td>
<td>Agent</td>
<td>ActionType</td>
<td>Event</td>
<td>Moment</td>
<td>Boolean</td>
<td>Fluent</td>
<td>Numeric</td>
</tr>
<tr>
<td></td>
<td>Object</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Action</td>
<td>Agent</td>
<td>ActionType</td>
<td>Event</td>
<td>Moment</td>
<td>Boolean</td>
<td>Fluent</td>
<td>Numeric</td>
<td></td>
</tr>
<tr>
<td></td>
<td>initially</td>
<td>Fluent</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>holds</td>
<td>Fluent</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>happens</td>
<td>Event</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>clipped</td>
<td>Moment</td>
<td>Fluent</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>f ::= initiates</td>
<td>Event</td>
<td>Fluent</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>terminates</td>
<td>Event</td>
<td>Fluent</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>prior</td>
<td>Moment</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>interval</td>
<td>Moment</td>
<td>Boolean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>Agent</td>
<td>Sell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>payoff</td>
<td>Agent</td>
<td>ActionType</td>
<td>Moment</td>
<td>Numeric</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t ::= x : S</td>
<td>c : S</td>
<td>f(1, 2, ..., i)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p ::= Boolean</td>
<td>¬φ</td>
<td>φ ∧ ψ</td>
<td>φ → ψ</td>
<td>φ ↔ ψ</td>
<td>∀x : S</td>
<td>φ</td>
<td>∃x : S</td>
<td>φ</td>
<td></td>
</tr>
<tr>
<td>P ::= B(a, t, φ)</td>
<td>D(a, t, holds(f, i))</td>
<td>I(a, t, happens(action(a, α), i))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>O(a, t, φ, happens(action(a, α), i))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

which when performed will lead to the agent simply believing that the value is something. There are costs associated with both the actions, and the agent would prefer knowing the value. More generally, if the agent believes that the cost of a full non-intuitive process is greater than the cost of an intuitive process by a threshold γ (which factors in the benefits of a deductive search), the agent would engage in an intuive process. Assuming that the agent knows some arithmetic, we can represent this model of intuition with this micro-theory:

\[
\begin{align*}
K(a, t, happens(action(a, α_{\text{deduce}}), t)) & \Rightarrow \\
\exists b K(a, t, valued(S(t + 1), b)) \\
K(a, t, happens(action(a, α_{\text{intuit}}), t)) & \Rightarrow \\
\exists b B(a, t, valued(S(t + 1), b)) \\
K(a, t, happens(action(a, α_{\text{intuit}}), t)) & \Rightarrow \\
\neg \exists b K(a, t, valued(S(t + 1), b)) \\
K(a, t, happens(action(a, α_{\text{intuit}}), t) \land (happens(action(a, α_{\text{deduce}}), t))) & , \\
k(a, t, happens(action(a, α_{\text{intuit}}), t) \lor \neg happens(action(a, α_{\text{deduce}}), t))) & , \\
B(a, t, \neg cost(α_{\text{deduce}}, t) \lor \neg cost(α_{\text{intuit}}) \land γ \Rightarrow (I(a, t, happens(action(a, α_{\text{deduce}}), t)))) & , \\
B(a, t, cost(α_{\text{deduce}}, t) \lor \neg cost(α_{\text{intuit}}) \land γ \Rightarrow (I(a, t, happens(action(a, α_{\text{intuit}}), t)))) &
\end{align*}
\]

2.4 On Logic and Intuition

It is not unfair to say that logic and intuition appear to be contra-positional: ‘intuitive’ seems to be the opposite of ‘rig-
Socioaffective intuition, concerning interpersonal relations and operating typically when a person seeks to understand a situation or another person.

2. “Applied intuition,” directed toward the solution of a problem or accomplishment of a task.

3. “Free intuition,” which arises in reaction neither to socioaffective stimulus nor to a specific search process, but rather involves “a feeling of foreboding concerning the future.”

It is doubtless the third type of intuition which has caused the diminution of intuition’s perceived value: how can something so intangible as a ‘premonition’ have any scientific value? Yet all three forms of intuition, including the second type, which is most immediately relevant to our work, have their bases in a structured, if unconscious, analysis and interpretation of perceived information. In our case, the perceived information is structured.

Because of its time-based nature, music is a good example where timely decisions have to be made, which immediately implies, on the Simon-Pollock view here affirmed, that intuition is needed. If time does not allow the deduction of a logical answer in the realm of music, humans must exploit mechanisms to make a quick “best guess” using a sub-logicial mechanism. Of course, we seek a system that does more than make a random guess by using an informed approach that at least approximates the correct answer.

Our agent, CAIRA, includes a logic- and music-calculus-based reasoning system, Handle, so it can “understand” basic concepts of music and use a hypothesis-driven approach to perform with other musicians. The agent uses logic to determine the current ensemble state from observing acoustical features in real time. CAIRA assesses the current ensemble state of an improvisatory music group, thereby addressing questions of who is playing a solo or whether it is likely that the improvisation will come to an end soon (see Table 4). For this purpose, the agent analyzes the relationships between the tension arcs of each musician, including the tension arc measured from CAIRAs own acoustic performance. The tension arc describes the current perceived tension of an improvisation, instrumentally estimated used acoustical features such and loudness and roughness [Braasch et al., 2011].

However, it often takes CAIRA too long to determine the current ensemble state using logic, so we conceived a process to simulate intuition, one that builds on a database of all currently known ensemble-state configurations. If an ensemble state is unknown for an occurring tension arc configuration (a vector with individual tension arcs for each musician), the process guesses the correct answer by taking into account the known states and making an informed guess using the eight nearest neighbors and distance weighting.

### 3 CAIRA Overview

Put briefly, CAIRA, the Creative Artificially-Intuitive and Reasoning Agent, the full architecture of which was presented at C3GI 2012, is a system able to provide guidance to ensembles, compose music, and perform in an improvisational fashion. A major goal of the CAIRA project is to develop an understanding of the relationship between logic and intuition within the scope of musical performances.
Instead of using a simple audio-to-MIDI converter, the agent uses standard techniques of Computational Auditory Scene Analysis (CASA), including pitch perception, tracking of rhythmical structures and timbre and texture recognition. This approach allows CAIRA to extract further parameters related to sonic textures and gestures in addition to traditional music parameters such as duration, pitch, and volume. This multi-level architecture enables CAIRA to process sound using bottom-up processes simulating intuitive listening and music performance skills as well as top-down processes in the form of logic-based reasoning. The low-level stages are characterized by a Hidden Markov Model (HMM) to recognize musical gestures and an evolutionary algorithm to create new material from memorized sound events. The evolutionary algorithm presents audio material processed from the input sound, which the agent trains itself on during a given session, or from audio material that has been learned by the agent in a prior live session. The material is analyzed using the HMM machine listening tools and CASA modules, restructured through the evolutionary algorithms and then presented in the context of what is being played live by the other musicians.

CAIRA consists of two standalone modules, FILTER and Handle.

3.1 FILTER
The artificially-intuitive listening and music performance processes of CAIRA are simulated using the Freely Improvising, Learning and Transforming Evolutionary Recombination system (FILTER) [van Nort et al., 2009; van Nort et al., 2010; van Nort et al., 2012], which uses a Hidden Markov Model (HMM) for sonic gesture recognition, and it utilizes Genetic Algorithms (GA) for the creation of sonic material. In the first step, the system extracts spectral and temporal sound features on a continuous basis and tracks onsets and offsets from a filtered version of the signal. The analyzed cues are processed through a set of parallel Hidden Markov Model (HMM)-based gesture recognizers. The recognizer determines a vector of probabilities in relation to a dictionary of reference gestures. The vector analysis is used to determine parameters related to maximum likelihood and confidence, and the data is then used to set the crossover, fitness, mutation, and evolution rate of the genetic algorithm, which acts on the parameter output space [van Nort et al., 2009].

3.2 Handle
Handle is a representative version of the logic- and intuition-based segments of the larger entity which is CAIRA, but nonetheless functions as a standalone creative machine conductor, as explored in prior work by Ellis and Bringsjord (see Figure 4). Whereas the prior version of Handle was designed to work with score-based, structured music with minimal deviation from a known form, its current incarnation is intended to explore the field of unstructured music in the form of improvised, free “fusion jazz.”

Now as formerly, Handle’s architecture consists of two major components, one running in MATLAB and the other in Common Lisp. The most notable change is that the MATLAB system is now the controlling client rather than the Lisp system as formerly. The MATLAB client interfaces with Filter and MaxMSP via a simple set of OpenAudio commands sent periodically via the network, which provide information on the current tension level of the three players as integers in the range 1-7. These values are used to look up the current ‘perceived state’ of the performance for Filter and the appropriate information returned. If, however, the state has not been encountered, a best guess is made, currently using a weighted nearest-neighbor heuristic, and the state is marked for formal calculation using an implementation of the music calculus in the Lisp-based SNARK automated theorem prover. Due to the length of time it takes to perform the operations, all logic calculations are done off-line between performances. An overview of the instantiation of the Handle architecture corresponding to the work reported on herein is given in Figure 5.

4 Demonstration
4.1 Method
In the demonstration, Handle, as the logical reasoning portion of CAIRA, accepts inputs from three audio sources: two of these correspond to two human players and the third is
the output from FILTER. Computational Audio Scene Analysis (CASA) is used on all three audio streams at various points during the piece to extract data on various features of each player’s performance. These features — tension$_R$, tension$_MIR$, dynamic, tempo, valence and activity — are mapped to an integer scale in the range 0–6 [Braasch, 2012], using combinations of 3-tuples of these features, information on the current state of the performance may be derived. Currently defined states and their definitions may be found in Table 1; possible states are solos for players A, B and C, low- and high-tension tutti, ending and an “uncertain” state. The latter state indicates there is insufficient information for Handle to supply meaningful data. When CASA information is supplied to Handle, it returns an integer value representing current state of the piece, using the definitions in Table 1. Based on this recommendation, FILTER then adjusts its own parameters to take the recommendation into account in its performance. At present, only tension$_B$ (as defined in [Braasch, 2012]) is used for state calculations.

As with human players, Handle may well have gaps in its knowledge when it is asked for its opinion on a state it has not encountered. In this case it will attempt an “intuitive leap,” using a predetermined heuristic to calculate a “best guess” response. If such a guess is made, it is stored in the knowledge-base together with a marker that it is a guess. At the end of the song, all states where a guess exists are passed to SNARK for formal proof, and the resultant values are stored permanently in the knowledge-base. Figures 6 and 7 show output from Handle operating over a short piece of music and the beginning of the resulting SNARK proof resulting from an unknown state, respectively.

### 4.2 Logic Subsystem

Handle operates first and foremost by logical deduction of state using axioms derived from predetermined rules. We use the rules given in Table 1 to construct premises in first-order logic by which one may prove each state. Numerical values received via the CASA system are mapped to a corresponding player-level value ($0 = A$, $1 = B$, etc.); these are used for comparisons during theorem-proving, replacing the ordinary arithmetic operators which SNARK cannot reason over.

#### Table 1: Ensemble states as determined by Tension$_B$, A, B, and C represent the calculated state of Musicians A, B, and CAIRA respectively. CAIRA may disregard the recommendation and respond differently. CAIRA must decide which state to prefer. States are ordered hierarchically such that overlap goes to the higher ranking state [Braasch, 2012].

<table>
<thead>
<tr>
<th>Ensemble States</th>
<th>Musician A</th>
<th>Musician B</th>
<th>CAIRA C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solo A</td>
<td>A &gt; B + 1</td>
<td>B + 1 &lt; A</td>
<td>C + 1 &lt; A*</td>
</tr>
<tr>
<td>Solo B</td>
<td>A + 1 &lt; B</td>
<td>B &gt; A + 1</td>
<td>C + 1 &lt; B*</td>
</tr>
<tr>
<td>Solo C</td>
<td>0 &lt; A &lt; 5</td>
<td>0 &lt; B &lt; 5</td>
<td>??</td>
</tr>
<tr>
<td>Low Level Tutti</td>
<td>0 &lt; A &lt; 5</td>
<td>0 &lt; B &lt; 5</td>
<td>??</td>
</tr>
<tr>
<td>High Level Tutti</td>
<td>A &gt; 4</td>
<td>B &gt; 4</td>
<td>C &gt; 4*</td>
</tr>
<tr>
<td>Ending</td>
<td>A = 0</td>
<td>B = 0</td>
<td>C = 0*</td>
</tr>
<tr>
<td>Uncertain</td>
<td>None</td>
<td>None</td>
<td>??</td>
</tr>
</tbody>
</table>

The following are some key examples of premises in FOL that we use to define the music states [Valerio, 2013], some of which are visible as SNARK axioms in Figure 7.

1. A person has a solo iff they have the highest level and there are no other players within one level.

   \[
   \forall p_1 : \text{Person} \ \exists p_1 : \text{Solo}(p_1) \leftrightarrow \exists l_1 : \text{MaxLevel}(p_1, l_1) \wedge \\
   \neg \exists l_2 : \text{Level}(l_2) \wedge \text{WithinOne}(l_1, l_2) \\
   \wedge (p_1 \neq p_2)
   \]
2. There is a low-level tutti iff every player is has a level greater than zero and less than five, and no player has a solo.

\[
Tutti(low) \iff \forall p: \text{Person } \exists \ell: \text{Level}
\left( \text{Level}(p, \ell) \land \text{LessThan}(1, f) \land \text{LessThan}(a, l) \land \neg \text{Solo}(p) \right)
\]

3. There is a high-level tutti iff every player has a level greater than four.

\[
Tutti(high) \iff \forall p: \text{Person } \exists \ell: \text{Level}(\text{Level}(p, \ell) \land \text{LessThan}(e, l))
\]

4. It is the end iff all players are at level zero.

\[
\text{End} \iff \forall p: \text{Person } (\text{Level}(p, a))
\]

By determining states logically in this fashion, the agent is guided to playing in a fashion that is expected on a social level. Given that the state determines certain weightings within the FILTER machine-learning aspect, however, the agent is still able to improvise freely within the expected range as a response to the other players. In effect, this is the same as a human player reasoning about what the other players are doing and what they intend, and reacting accordingly. Similarly, when the agent is first initialized, it has no knowledge about what combinations of inputs relate to which state and must also learn by reasoning about new inputs over time.

### 4.3 Approaching Intuition

Handle is a logical system that uses axioms provided and computed in SNARK to derive state values matching to combinations of 3-tuples of musical features. However, again, in cases where precomputed information is unavailable, Handle has been designed to make an “intuitive leap,” by using one of a number of possible heuristics to guide its judgment.

We note at this point that for the given three-player scenario we may represent each state as a color corresponding to the most extreme coordinates for a feature 3-tuple. For example, a solo for player A corresponds to red (6, 0, 0 \rightarrow 255, 0, 0), Solo B is 0, 6, 0, which encodes to green, and Solo C is 0, 0, 6, blue. Low-tension tutti is represented by grey (calculated by finding the centroid of the state), high-tension tutti is white, end is black, and uncertainty is determined by the coordinate values themselves. After computing all state combinations, such a representation appears as in Figure 8.

The current approach for resolving an unknown state at runtime is to use a nearest-neighbor algorithm. Given the (safe) assumption that the states for all 8 corners of the state cube are absolutely known ab initio, the value of any unknown state may be derived by taking the average of the sum of the distance vector between it and each (fully known) corner. As more and more states are derived and the knowledge base becomes progressively more complete, however, we would prefer to use more accurate information; that is, state values which are more proximate to the unknown state. These 8 cubes, nearest to the unknown state in diagonal distance, are determined using a simple vector approach and the mean of their values returned. Thus, given an unknown state σ with coordinates (α, α, α) and (x), (y), (z) vectors of the coordinates of all known states, the 8 nearest neighbors of σ can be found as follows:

\[
\begin{align*}
\text{index}_1 &= \text{find}(s[i] \geq \alpha_x \land y[i] \geq \alpha_y \land z[i] \geq \alpha_z) \\
\text{index}_2 &= \text{find}(s[i] \geq \alpha_x \land y[i] \leq \alpha_y \land z[i] \geq \alpha_z) \\
\text{index}_3 &= \text{find}(s[i] \geq \alpha_x \land y[i] \leq \alpha_y \land z[i] \leq \alpha_z) \\
\text{index}_4 &= \text{find}(s[i] \geq \alpha_x \land y[i] \leq \alpha_y \land z[i] \leq \alpha_z) \\
\text{index}_5 &= \text{find}(s[i] \leq \alpha_x \land y[i] \geq \alpha_y \land z[i] \geq \alpha_z) \\
\text{index}_6 &= \text{find}(s[i] \leq \alpha_x \land y[i] \geq \alpha_y \land z[i] \leq \alpha_z) \\
\text{index}_7 &= \text{find}(s[i] \leq \alpha_x \land y[i] \leq \alpha_y \land z[i] \leq \alpha_z) \\
\text{index}_8 &= \text{find}(s[i] \leq \alpha_x \land y[i] \leq \alpha_y \land z[i] \leq \alpha_z)
\end{align*}
\]

Admittedly simplistic, this approach nonetheless conforms to our definition of ‘intuition’ given in section 4. In order to determine the efficacy of this approach, a simulation was run of over 622 sessions to determine the percentage error of the nearest-neighbor approach by picking 10 random player inputs each session. It can be seen that, as new states were accumulated in the knowledge base, the accuracy of the nearest neighbor method improves (Figure 10). A comparator is also provided, based on an approach which guesses random on uncomputed states.

We acknowledge that we may stand accused of not properly addressing, or exploring, the concept of ‘intuition’ in this iteration of Handle. We do not deny this. However, in our own defense, we would reply by stating that ‘intuition’, as it is commonly understood, requires a much larger dataset than the one we are currently using, one that in all likelihood contains a certain amount of conflicting data, together with an appropriate means of accessing and using those data. The current incarnation of Handle functions using only one musical feature, tension, yielding a dataset containing a paltry 7\times7\times7=343 states; as a consequence, ‘intuitive leaps’ will inevitably be barely distinguishable from mere arithmetic computation. But when we do have the additional information, the problem inevitably arises of how we manage it.

One of the most interesting effects to arise from this iteration of Handle is the concept of the state cube, shown above in Figure 8. Originally developed by Valerio based on work by Ellis, it has considerable implications for computation using...
can model this reasoning — and build agents that can perform such complex reasoning in real time. We now present one embryonic example of this, by examining a pattern of reasoning carried out by a conductor.

The conductor could be unsure whether a performance should have a certain attribute; for example, proportional tempo between movements. If the conductor then comes to believe, via some process, that the composer believed that the performance’s having this attribute would result in a pleasing experience for the listener, then the conductor may start believing likewise. Naturally, this “epistemic chain” would result in performances that have proportional tempo.

The reasoning pattern above can be computationally realized using the following proof in $\mathcal{DCEC}^\ast$. We first present the necessary premises (Performance and Emotion are subsorts of Action and Fluent respectively):

\[ A_1 \forall a, f, t, f', D(h, t, \text{holds}(f, t)) \land B(h, t, \text{initiates}(\text{action}(h^\ast, a), f, f')) \Rightarrow \text{I}(h, t, \text{happens}(\text{action}(h^\ast, a), t + 1)) \]

\[ A_2 \forall \rho : \text{Performance, } \alpha : \text{Emotion}, t, t'. \]

\[ B(h, t, B(c, t, \text{initiates}(p, \text{holds}(\text{feels}(\alpha, \text{listener}), t'))))) \land \neg B(c, t, \text{initiates}(\neg p, \text{holds}(\text{feels}(\alpha, \text{listener}), t'))))) \Rightarrow B(h, t + 1, \text{initiates}(p, \text{holds}(\text{feels}(\alpha, \text{listener}), t'))) \]

\[ A_3 \forall t, t' : B(c, t, \text{initiates}(\text{action}(h^\ast, \text{perform}(\text{tempo}_{\text{prop}})), t')) \land \neg B(c, t, \neg \text{initiates}(\text{action}(h^\ast, \text{perform}(\text{tempo}_{\text{prop}})), t''))) \]

\[ A_4 \forall t, t' : D(h, t, \text{holds}(\text{feels}(\text{pleased, listener}), t')) \]

The informal English translations of the premises are given below:

\[ A_1 \] If the conductor desires something and believes a certain action will accomplish it, the conductor intends to perform that action.

\[ A_2 \] For all performances $p$ and emotional states $e$, if the conductor does not know whether $p$ can lead to state $e$ in the listener, but believes that the composer believes $p$ can lead to emotion $e$ in the listener, then the conductor will then come to believe the same.

\[ A_3 \] The conductor himself is ambivalent about the above statement that movements with proportional tempos will lead to the listener being pleased.

\[ A_4 \] The conductor desires that the listener be pleased.

Assume we have constant symbols with $t_1 < t_2 < t_3 < \ldots$ and that if $t \equiv t_i$ then $t_{i+1} \equiv t_i + 1$, and that we also have the following along with the above five premises (where the composer communicates with the conductor in some fashion):

\[ \forall S \left( c, h, t_1, \forall a : \text{initiates}(\text{action}(a, \text{perform}(\text{tempo}_{\text{prop}})), t')) \right) \]

\[ ^1 \text{This is our “logification” of actual reasoning performed by the famous conductor Marriner; see } [\text{Anderson, 1985}]. \]
Using universal elimination, \( R_{12} \), and universal introduction with the above, we obtain the following:

\[
\phi_t = \\
\forall' B \left( h, n, B \left( c, n_1, \forall a : \text{initiates} \left( \text{action}(a, \text{perform}(\text{tempo}_{\text{query}})), t' \right) \right) \right)
\]

Let

\[
\phi_t^{\text{surface}} = \\
\forall' B(h, t_2, \text{initiates}(\text{action}(h', \text{perform}(\text{tempo}_{\text{query}})), \text{feels}(\text{pleased}, \text{listener}), t'))
\]

Finally, we have the following proof that the conductor will try to bring about a performance with proportional tempo.

\[
\frac{\gamma t_1 \phi_t}{\gamma a \psi} \implies t(h, t_2, \text{happens}(\text{action}(h', \text{perform}(\text{tempo}_{\text{query}}), t_2)))
\]

6 Next Steps

While the current research has had some genuinely interesting outcomes, as always it creates at least as many questions as it answers, for example:

1. Further development of axioms and rules in the music calculus is required.

2. We will explore the possibilities of bi- and multi-variate analyses of musical performances, together with different ways of encoding the combined state data. Metrics we have considered include tempo or dynamic, at given moments or across a period, and another form of tension derived from the ‘emotional’ strength of a piece of music at a given time.

3. It would be useful and interesting to explore different heuristics for the nearest-neighbor analysis generally. Are there calculations which work well in some cases but not others? What happens if the heuristic is biased toward states nearer the query state and against those further away? How good is the estimate compared to the actual calculated state, and is there any form of pattern to the wrong guesses?

Acknowledgements: This project is made possible by generous sponsorship from both NSF (grant no. 1002851) & the John Templeton Foundation.

References


How Models of Creativity and Analogy Need to Answer the Tailorability Concern*

John Licato, Selmer Bringsjord, Naveen Sundar Govindarajulu
{licatj,selmer,govinn}@rpi.edu
Rensselaer AI and Reasoning (RAIR) Lab
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA

Abstract

Analogy is a major component of human creativity. Tasks from the ability to generate new stories to the ability to create new and insightful mathematical theorems can be shown to at least partially be explainable in terms of analogical processes. Artificial creativity and AGI systems, then, require powerful analogical sub-systems—or so we will soon briefly argue. It quickly becomes obvious that a roadblock to such a use for analogical systems is a common critique that currently applies to every one in existence: the so-called ‘Tailorability Concern’ (TC). Unfortunately, TC currently lacks a canonical formalization, and as a result the precise conditions that must be satisfied by an analogical system intended to answer TC are unclear. We remedy this problem by developing a still-informal but clear formulation of what it means to successfully answer TC, and offer guidelines for analogical systems that hope to progress further toward AGI.

1 Introduction

Creativity and the ability to reason analogically have a strong relationship [Finke, 1990, Finke et al., 1992, Holyoak and Thagard, 1995, Gentner et al., 1997]. For example, many of the greatest lines in literature make use of familiar analogical processes: “A rose by any other name would smell as sweet” only makes sense if that rose and its name are understood to correspond to Romeo and his family name (analogical mapping). The subsequent analogical inference is that just as the nature of the rose is independent of its name, so is Romeo’s nature independent of his.

Even if one believes that overuse of analogy can be harmful to creative thought [Riesbeck and Schank, 1989], many researchers argue that the ability to determine analogical similarity is important at least for combinatorial creativity, which seems to be very easy for humans but very difficult for AI systems [Boden, 1995, Boden, 2009]. Furthermore, the importance of analogy is of course not limited to creativity, as analogical ability has been identified as an indispensable component of artificial general intelligence as well [Bringsjord and Licato, 2012, Hofstadter, 2001].

With the above in mind, it makes sense to develop models of analogy, both computational and theoretical. Much work has been done in this direction; a few implementations include SME [Falkenhainer et al., 1989], LISA [Hummel and Holyoak, 2003], HDT [Gust et al., 2006], and recently our own META-R [Licato et al., 2013]. On the surface, it seems that the current generation of analogical systems sufficiently capture and explain all of the phenomena commonly associated with analogical reasoning, and that they will eventually reach levels characteristic of human cognition. It may well be the case that the most important principles underlying the nature of analogy have been expressed. But a serious objection has been raised recently which, as will be argued, should be the primary focus of analogical researchers over the next few years — at least if any significant further progress is to be made in the direction of creativity and AGI.

The objection is raised by Gentner and Forbus (2011). They call it the ‘tailorability concern’ (TC), and the objection echoes a common criticism of cognitive systems in general: that they operate on toy examples manually constructed in such a way as to guarantee the desired solution. However, though this concern may have been stated in many forms throughout the years [Mitchell and Hofstadter, 1995], it lacks, to our knowledge, a formulation clear enough to anchor productive scientific discussion. And this ambiguity in turn negatively impacts not only the relevant science, but AI engineering as well: absent a definition of TC, it is difficult to understand precisely what an analogical system must do in order to successfully answer TC. In the remainder, we take steps toward addressing this problem as it applies to analogical systems.

2 The Tailorability Concern

A frequently appearing criticism of cognitive systems in general is that they are only applied to manually constructed ‘toy examples’, a problem many researchers in the field themselves acknowledge. Gentner and Forbus (2011) referred to it as the ‘tailorability concern’ (TC): “that is, that (whether knowingly or not) the researchers have encoded the items in such a way as to give them the desired results” [Gentner and Forbus, 2011].

---

*With profound gratitude, the support of both the John Templeton Foundation and AFOSR is acknowledged.
The field of case-based reasoning (CBR), which overlaps quite heavily with that of analogical reasoning (AR), also deals with some of the issues raised by the TC. There are differing opinions on what features distinguish the CBR and AR approaches (see [Aamodt and Plaza, 1994] [Burstein, 1989] [Mantaras et al., 2005]), but two common themes are that CBR tends to deal with source and target cases that come from the same domain, and cases are selected and adapted with some clear pragmatic goal in mind. AR approaches, on the other hand, try to be more generally applicable across different domains, and tend to focus more on the mapping process that actually determines analogical similarity. CBR approaches, then, deal with the TC by trading generality for effectiveness, so that a program designed to work well in one domain (medical diagnosis, for example, is a popular field) may not work so well in another without a significant amount of human assistance.

Unfortunately, the CBR approach of restricting generality does not sufficiently answer the TC. Analogy research can be seen as centering around fundamental questions, one of them being: How can we find good analogies? The TC is especially problematic because it forces analogy researchers to prove that their theoretical process is the answer to this question, and although it can be promising to see that some particular approach produces good analogies in some limited domain, no approach can constitute a completely satisfying answer to this question unless it is versatile enough to perform well in many domains.

Any system that can answer the challenge of the TC will instantly distinguish itself from every other analogical system currently in existence, since to our knowledge the only one that has been able to do this with some degree of success is the SME-based family of systems [Gentner and Forbus, 2011] [Lovett et al., 2010] [Forbus et al., 2009]. For this reason, we should clarify what it means to answer this challenge and discuss why it is such a non-trivial feat.

2.1 Answering the TC

Gentner and Forbus (2011) suggest that there are two possible ways to answer the TC. One applies to visual domains, and involves using automatic encodings of visual representations. The other more generally applicable direction involves two key features: First, the use of pre-existing databases; second, an automated or semi-automated parsing process that goes from input text to a sufficiently rich semantic representation. A first attempt at a precise statement of what it means to answer the TC is as follows:

$$\text{TCA}_1$$ A computational system of analogy answers the TC if, given no more than a pre-existing database and an unparsed input text, it is able to consistently produce good analogies across many domains.

At least one general intent behind $$\text{TCA}_1$$ is clear: it attempts to place emphasis on the filtering process (whose job is, as we said, to either select some subset of available source analogs from a large database and recommend only some of them for the more computationally expensive step of analogical matching, or to automatically construct structured representations from sensory data). By removing the reliance on

Figure 1: One version of the popular “heat flow” example from Falkenhainer et al. (1989) used to test many analogical systems.

Of course, nothing is wrong with toy examples per se. They can be extremely useful in demonstrating key concepts, helping to illustrate particular qualitative strengths or weaknesses of computational models, or helping to get a new model off the ground. Indeed, the present authors plead guilty to using toy examples this way; properly done, carefully chosen cases can be very useful as demonstrations-of-concept. But we should be careful not to treat these as the final proof of a system’s worth, since in most of these examples it is not clear that the principles used to solve them generalize to other problems, and can be used to mechanically find useful solutions just as effectively without human assistance.

For example, the well-known water-flow/heat-flow analogy (Figure 1) has been used as a demonstration of many models of analogy [Falkenhainer et al., 1989] [Holyoak and Thagard, 1989] [Wilson et al., 2001] [Krummack et al., 2007]. But little to nothing is written about how the structural representations used in examples such as these are acquired in the first place. One approach is to model the acquisition of structured representations through sensory data (e.g., see [Doumas et al., 2008]), and another is to presume the existence of a large database of already-structured data (such as that to which a neurobiologically normal adult might be expected to have access), and some sort of filtering process that ensures that from this database, proper representations are selected and any unnecessary data that would produce incorrect matching results are excluded. Yet even when such filtering processes are proposed, they are not put to the test and proven to perform well with a large source of data whose size approaches that of a child, much less an adult. The TC right-fully attempts to refocus efforts on these filtering processes, by requiring that they demonstrate the ability to produce clean source and target analogs as required by the analogical map-
human intervention. TC ensures that the filtering is not done manually in such a way that guarantees desired results. However, in order to truly answer the TC in a satisfactory way, we must be precise about its purpose and motivations: What concerns are behind the TC in the first place? Furthermore, TCA₁ is hopelessly vague: the words ‘unparsed’ and ‘good,’ if left open to interpretation, make it too easy for anyone to prematurely claim victory over TC. Also: Why is it important that the database be pre-existing? What degree of separation must there be between the creators of the database and the designers of the analogical system? For example, does the database’s underlying knowledge-representation philosophy need to overlap with that of the analogical system?

What is a ‘good’ analogy?

Though the widely influential Structure-Mapping Theory [Gentner, 1983] offers the systematicity principle and the one-to-one constraint as indicative features of a good analogy, it does not provide a clear quantitative measure of analogy quality. SME evaluates match scores by combining the scores of the evidence provided by its match rules; this allows for a comparison between different matches of the same problem. But the resulting match score is not normalized, and as a result match quality between different problems cannot be compared [Forbus and Gentner, 1989]. Other analogical models do not help much in this regard. Holyoak and Thagard’s (1989) Multiconstraint Theory, for example, introduces additional criteria to evaluate what makes an analogical match a good one, making cross-domain analogical match quality more difficult to assess.

This is especially problematic when considering TC. Given a set of predefined problems and desired answers for all of them, the good analogy is simply the one that performs as intended. But when applied to problem sets where the best analogy, or even the existence of a good analogy, is not clear even to the persons running the experiment, do the guidelines currently available still apply?

Paul Bartha (2010) offers a comprehensive theory of analogy that can be useful here. His goal is to produce a normative account of analogy, and in the process he sheds light on the proper role of analogy within a larger context of reasoning. His Articulation Model is based on the idea that there are two features common to all good analogical arguments: a prior association (a relationship in the source analog that is transferred to the target), and a potential for generalization. But perhaps most relevant to our present discussion is his claim that by virtue of the analogical argument, the resulting hypothesis inferred through an analogical argument contains no more than prima facie plausibility, which can be understood as something like a suggestion that the inferred hypothesis is worth exploring further. If there is an independent reason to reject the hypothesis, such as a deductive argument showing it leads to a contradiction, or contrary empirical evidence, then it can be abandoned.

The idea that the proper role of an analogical argument is to do no more than provide prima facie plausible hypotheses [Owen, 1990] [Norton, Forthcoming] suggests that the relationship between an analogy’s match quality and its tendency to produce hypotheses which can be independently verified may not be as simple as it might seem. In the end, a model of analogy is a good one only if it produces good analogies, and an analogy is a good one only if it produces plausible hypotheses.[2]

This complicates things further. In toy examples, the information available to the analogical matcher is very limited; a hypothesis in these cases is plausible if the information available to the analogical matcher does not provide a reason to reject that hypothesis. But if we are suggesting that the information available to the analogical system (or more specifically, the filtering process) is actually a large database conceivably encompassing, say, all of the knowledge on the World Wide Web, then how are we to realistically evaluate the system’s performance?

In light of these complications, we propose the following alterations of TCA₁. If the word ‘good’ is replaced with ‘useful,’ then we connote an evaluation method that is not based in the particular metrics as defined by the analogical theory itself (which can be controversial), but rather based in an intuitive notion that can and should be evaluated independently of the model. In other words, researchers might disagree on how to measure an analogical match, but whether the resulting analogically inferred hypothesis is useful can be evaluated without any knowledge of the analogical matcher used. Of course, we do not claim that the word ‘useful’ is completely unambiguous, but rather open to interpretation, as it might be domain dependent. Later in this paper, a better replacement for the word ‘useful’ will be suggested. For now, the aim behind this move is to divorce the metric which determines the quality of an analogical match’s results (which may be very domain-dependent) from the theory-specific metric that the matcher is specifically designed to optimize. That (deceptively) small change gives us TCA₂:

TCA₂ A computational system of analogy answers the TC if, given no more than a pre-existing database and an unparsed input text, it is able to consistently produce useful analogies across many domains.

What are acceptable databases?

In TCA₂, it is clear that the information available in the database used is a limiting factor in how knowledgeable the inferences produced by the analogical system can be. The suggestion phrased by Gentner and Forbus (2011) as “pre-existing databases” requires more clarification. The implication (at least as we interpret it) is that the dataset and the structures within cannot have been constructed for the purpose of solving the particular toy examples that are of interest. Otherwise this introduces bias and tailorable concerns, in spite of the best intentions of the designers. Two issues immediately come to mind. First, what is the proper level of separation between the database and the analogical system? Secondly, how large does the database have to be?

[2]We might leave room here to exclude models of analogy that have psychological or neurological plausibility as their primary end goals. In these cases, it might be the goal of the model to replicate poor analogical reasoning as well, if it matches human performance. But it is our assumption (at least in the present inquiry) that the ultimate goal of AGI research is not to model poor human reasoning.
The first question is at least partially answered by considering the level of representational agreement between the database and the analogical system. For example, if the database is a purely distributed one with no localist concepts whatsoever (which is, we acknowledge, an unlikely possibility), and the analogical system is one that uses only localist, explicitly structured data, then a significant level of work will be needed to first extract the information from the database and put it into the form that the analogical reasoner requires (this can be considered a re-representation step [Chi et al., 1981; Yan et al., 2003; Krumnack et al., 2008]). The choice of database becomes important for this reason, and if no database exists that does not require a high level of re-representation, then it suggests a problem: Consider that although proponents of localist, distributed, and hybrid representation styles make claims all the time about the scalability of their assumptions of knowledge representation, the researchers who have to design and work with large semantic databases actually have to put these ideas to the test. If the state-of-the-art research in the database field has significant difficulty with some style of representation as required by an analogical matcher, then perhaps that matcher needs to carefully consider its assumptions about the nature of knowledge representation, or else be able to easily extract the necessary data. The choice of database is therefore a sensitive one.

It may help to go back to the criticisms which may have motivated this requirement in the first place. Hofstadter’s group criticized SME by saying that “the analogy is already effectively given in the representations” [Mitchell and Hofstadter, 1995]. The information provided to the matcher is selected in such a way that it does not include a significant amount of extra data that would lead to false analogies, and the choice of predicates and objects is done in such a way that presupposes the desired matching. This is a criticism composed of two points: the amount of information provided as input (which is not too little, or too much, but just right), and the nature of the information (the corresponding predicates on both sides just happen to have the right amount of arguments in the necessary order). The system is not robust enough to produce good analogies when given the same input data in a variety of formats.

Two possible changes can begin to answer these critiques. One is to require that the databases are large enough, and the input minimal enough, to introduce a significant amount of false matches that would confuse less robust analogical matchers. Critics claiming that the size of the input is too carefully chosen could be pointed to the large search space that an analogical matcher must plumb given the small input problem size. The larger the search space, the more impressive the ability of the analogical matcher to select only a few potentially relevant source analogs. Furthermore, an inability to demonstrate scalability to large datasets weakens any architectural’s claim to psychological plausibility: if the architecture can’t handle a dataset large enough to produce non-trivial answers, how can it be an accurate model of a process used by human-level reasoners?

Secondly, we could require robust and consistent performance on a variety of input forms. For example, in the heat-flow problem (Figure 3) Mitchell and Hofstadter (1995) note that there are many possible ways to structure the input: heat could be described as an object, or as an attribute of coffee, or heat flow could be a relation with three rather than four arguments [Mitchell and Hofstadter, 1995]. Consistent performance across various input forms puts more pressure on the analogical matcher’s re-representation algorithm(s), rather than relying on a separate NLP module. This also allows for a leveling of the playing field across different systems: In order to show an example adheres to this requirement, a localist, structured analogical system would have to demonstrate two things with regard to that particular example:

- Across multiple, more or less equivalent structural representations of the same input data, and a wide variety of domains, the matcher still produces the desired results.
- The desired results are still produced when the input is minimal: meaning any redundant information or structural constructs which might be identified by critics as being used only to aid the matcher can be removed.

Of course, if a system did happen to require no more than unstructured, natural-language descriptions as input, or direct sensory data from visual representations, it would satisfy both of these conditions. This allows our criteria to encompass the alternate route to answering the TC mentioned by Gentner and Forbus (2011)—a route which seeks to answer TC by not necessarily having a large database, but by having one that at least attempts to directly construct structured representations from low-level sensory or natural-language data. The generality of these conditions allows us to claim the converse of TCA2, leading us to our next iteration:

TCA3 A computational system of analogy answers the TC if and only if given no more than either
- unstructured textural and/or visual data, or
- a large, pre-existing database,
and minimal input, it is able to consistently produce useful analogies and demonstrate stability through a variety of input forms and domains.

One might be satisfied with this set of criteria, which draws its strength from its lack of commitment to any particular theory of analogy, and its emphasis on large sets of non-tailored input. But TCA3 is undeniably ambiguous, and may not be focused enough to guide any research program. We encourage the reader to take TCA3 and develop it further, but first, to close out this paper we will take some steps of our own to reduce some of its weaknesses.

---

1We do not mean here to say that what works best for large artificial databases is the same as what is employed by the human brain. But if a researcher declares that the structure of human knowledge has certain properties, and large datasets cannot be created that do not have those properties for reasons of scalability, then it should be at least a weak hint that perhaps the assumption of those properties is not practicable.

2This is a common criticism of Hummel and Holyoak’s LISA system; see Gentner and Forbus, 2011. 

---
Strengthening the TC with Psychometric AI

We will make two important moves to sharpen TCA$_3$. One, we turn to Psychometric AI (PAI) \cite{Bringsjord2011,Bringsjord2003}, according to which, in a nutshell, commendable AI systems are those that demonstrate prowess on tests of various mental abilities from psychometrics. Our second move is to imbue analogy-generation systems within broader AI problem-solving systems that make use of additional forms of fundamental reasoning in human-level intelligence; e.g., deduction. In particular, we place TC within the context of the integration of analogical reasoning with deduction, which we dub analogico-deductive reasoning (ADR). An ADR system does use analogy generation, but analogies are used to guide solutions that can be rigorously verified by proof. The architecture-sketch of an ADR system that accords with our pair of moves is shown in Figure 2. While we don’t have the space to provide details here, this system receives puzzles that are part linguistic and part visual in nature (e.g., so-called seating puzzles), and harnesses not only our own analogical system (META-R), and not only ATP technology, but visual theorem-proving; for formal details see \cite{Arkoudas2009}. This architecture-sketch is inspired by, but abstracts and extends beyond, AI systems able to solve analogy-relevant problems. One such inspiring system is by Lovett, Forbus, and Usher \cite{Lovett2010}; it can solve items on the Raven’s Progressive Matrices. But it cannot for example prove that its answers are correct, which is part of what the architecture-sketch in Figure 2 demands.

Given the context we have now set, we can articulate a new biconditional:

$\text{TC}_A \iff \text{A for analogy generation answers the TC if and only if, given as input no more than either unstructured textual and/or visual data, or a vast, pre-existing database not significantly pre-engineered ahead of time by humans for any particular tests of } A.$

is — in keeping with co-called Psychometric AI — able to consistently generate analogies that enable $A$ to perform provably well on precisely defined tests of cognitive ability and skill.

To comment briefly on $\text{TC}_A$, first note that we remove the explicit requirement that the ability to find useful analogies be stable across a variety of input forms and domains. This is subsumed by the requirement of good performance on precisely defined tests; it is assumed that a sufficiently difficult psychometric test would provide questions that are both varied in their form (e.g., word problems, puzzle solving, story comprehension) and in their domains. The implicit requirement of domain variety rules out the possibility of an artificial reasoning agent that can only process, for example, certain types of math problems, as an acceptable answer to the TC.

Some might also see $\text{TC}_A$ and the use of PAI as too restrictive in that it relies too heavily on problem solving and not enough on either creative thinking or the kind of everyday analogical thinking that may not be goal-oriented in nature. PAI, however, provides a tool for measuring those abilities that, at least at the surface, don’t rely on directed problem solving, such as reading comprehension. Additionally, it is difficult to imagine that any research program in AGI would be able to demonstrate clear progress without showing increased performance in an ability that can be measured according to some psychometric test. Non-goal-oriented analogical reasoning is a good example of this principle: If the cognitive processes underlying normal analogical reasoning when it is non-goal-oriented (as in everyday reasoning) and when it is goal-oriented (as during psychometric testing) are largely the same, then an artificial system capable of performing the latter has a strong claim to performing the former. A system that only has sporadic performance on psychometrically-measurable tasks is difficult to defend as generally intelligent.

One might ask: Can a system do well on psychometric tests and still be subject to claims of tailorability? The answer, if the requirements in $\text{TC}_A$ are not met, is yes. PAI is not meant to be a replacement for the input format and large database requirements we have been developing in this paper; rather, it is only one possible sharpening of the ambiguous concepts in $\text{TC}_A$. Other possibilities may exist, but we do not at present know of any satisfying alternatives.

3 Conclusion and Future Work

The concerns reflected by TC have been more or less echoed in many criticisms of analogical systems in the past 30 years or so, but little to no discussion has attempted to precisely state the conditions under which an analogical system would no longer be subject to this common complaint. We humbly...
acknowledge the limitations of our formulation here, and remind the reader that even a system with a strong claim to meeting the conditions of $\text{TCA}_4$ (e.g., an implementation of Figure 2) may not be completely immune to TC-based criticisms. But it seems that such criticisms would be somewhat difficult to justify. We hope that this paper inspires further discussion, and invite researchers to pick up from and sharpen $\text{TCA}_4$, or to steer away from Psychometric AI and therefore develop $\text{TCA}_3$ in a different direction.

Those who strive for something deeper than a test-based anchor for a response to TC will presumably be more sanguine than Bringsjord about the ultimate reach of computing machines. Bringsjord holds that it is impossible for any mere computing machine to originate anything [Bringsjord et al., 2001], or even to genuinely understanding anything [Bringsjord and Noel, 2002]. But he holds on the other hand that for any determinate behavioral test, a computing machine can be programmed to pass it—even if the test is one which, if passed, suggests to most human observers that the passing machine is not simply the product of human intelligence, harnessed [Bringsjord, 1995]. In short, the turn to $\text{TCA}_4$ as a way of fleshing out $\text{TCA}_3$ reflects deep pessimism about computing machines, and specifically a deep affirmation of the penetrating and profound nature of TC. Licato, in contrast, is more optimistic about the potential of analagical-hybrid systems, but is not prepared to take any philosophical positions on whether a system that passes TC demonstrates true understanding or creativity.

What about AI work devoted to building creative systems on the strength of analogical reasoning? Our work continues along this path. We are currently investigating the potential of using our analogical matching system META-R, which is designed to allow for a flexible application of heuristics across a wide variety of domains, with the community-built dataset Freebase [Bollacker et al., 2008]. We are also investigating the use of META-R in modeling the automatic generation of logical theorems and high-level proofs from semi-formal and formal domains [Licato et al., 2013; Govindaraju et al., 2013]. Future work in this direction includes the investigation of Analogico-Deductive Moral Reasoning (ADMR) using a hybrid of CLARION [Sun, 2002] and LISA [Hummel and Holyoak, 2003]. And finally, we are seeking to leverage analogical reasoning in order to engineer systems capable of automatic programming.

References

[Aamodt and Plaza, 1994] Agnar Aamodt and Enric Plaza. Case-based reasoning: Foundational issues, methodolog-


Creative Concept Invention and Representation in an Interactive Child-Robot Healthy Snack Design Game

Andrew B. Williams, Adam Stroud, Joshua Panka, John C. Williams
Humanoid Engineering and Intelligent Robotics Laboratory
Marquette University, Milwaukee, WI USA
{andrew.williams, adam.stroud, joshua.panka, john.c.williams}@marquette.edu

Abstract
In this paper, we describe how newly created metaphor concepts will be invented and represented by a humanoid robot. The humanoid robot will invent these new concepts during interaction with a child in a pretense play scenario. We demonstrate the feasibility of this concept invention and representation approach by evaluating a knowledge-based metaphor discovery program using an open source reasoning engine.

1 Introduction
The ability for a human to use her creativity to invent new concepts or artifacts is a hallmark for human intelligence. In order for a humanoid robot to attain human-like intelligence, it must first be able to represent and incorporate creative new concepts into its knowledge base and be able to make new inferences using this new knowledge. We propose using PyClips, a python language interface to the CLIPS knowledge-based programming language [Giarratano and Riley, 1998], as a tool to represent and reason about these newly discovered creative semantic concepts by a robot.

In [Williams, 2013] we described a newly proposed approach towards investigating how social humanoid robots can learn creative conceptualizations through interaction with children in metaphor-guided pretense play using a componential creativity framework. We describe this metaphor-guided pretense play approach and then illustrate it by describing a social robot interaction pretense play scenario between a child and a humanoid robot. The rest of this paper is organized as follows. First we review the creativity framework that we combine with our metaphor-guided pretend play creativity approach [Williams, 2013]. We then describe how our humanoid robot will learn new creative concepts using metaphors learned in a human-robot interaction scenario. Then we describe how CLIPS will be used to represent and reason about new creatively learned concepts.

1.1 Creativity Framework for Concept Invention
Concept invention is the process of discovering new concepts that may require new interpretations and mappings of of a given situation [Reffat, 2002]. Creativity is closely related to concept invention and is the quality that is marked by the ability or power to produce through imaginative skill. Creativity has been defined as the ability to generate novel, and valuable, ideas (e.g. concepts, theories, interpretations, or stories) [Boden, 2009]. We use a componential framework for creativity that is based on three factors for creativity: domain-relevant skills, creativity relevant skills, and task motivation [Amabile, 1983]. Amabile [1983] presents a schematic for the creative process based on these skills and motivation that includes:

1. Problem or task representation
2. Preparation
3. Response generation
4. Response validation
5. Outcome

We incorporate this framework with metaphor-guided pretend play approach for creative concept invention and discovery.

1.2 Concept Invention through Metaphor Learning and Pretense Play
In this section we describe how metaphor is related to creativity, or concept invention. According to Indurkhya [1992], a “metaphor is an unconventional way of describing (or representing) an object, event, or situation (real or imagined) as another object, event or situation”. [Genter et al., 2001] describe the critical steps for using metaphors for problem-solving. One must extract a variety of unfamiliar concepts from remote domains where possible relationships with the current task may not be initially apparent. Then there must be a mapping of high-level or deep relationships between the metaphor concept and the problem. Performing generalization and abstraction techniques may make this matching possible. Next, secondary relationships may be discarded leaving only structural correspondence between the metaphor source and the problem. Finally, the structural matches, or correspondences, associated with the metaphor source are transferred and applied to the problem, which
leads to a novel solution. Using metaphors is an example of a creativity-relevant skill of generating hypotheses that can lead to set-breaking and novel ideas. These metaphors can be used as heuristics to organize problem solving or design thinking to solve loosely defined design problems [Rowe, 1987; Antoniades 1992].

2 Concept Invention in an Interactive Healthy Snack Food Design Game

We propose using the context of a healthy snack food design game to demonstrate how humanoid robots can invent new concepts through interaction with children. [Morris et al., 2012] have developed a system for generating novel recipes and use their system, PIERRE, to explore theoretical concepts in computational creativity. They show that a single inspiring set can be separated into two sets used for generation and evaluation to result in greater novelty. Other artificially intelligent food recipe generators include CHEF [Hammond, 1986] which used case-based reasoning to plan a recipe and Julia [Hinrichs, 1992] which also used case-based reasoning but to plan out a meal.

Children can create meaning from their social interactions collectively through pretense play [Janes, 2002]. They can creatively take on imaginary roles, act out stories, and change toys into tools or instructions. In metaphor-guided pretense play [Williams, 2013], the child is given the pretense play role of head chef with a set of healthy food items such as grapes, bananas, celery, soy butter, and raisins (Figure 1). The humanoid robot “customer” presents a “seed” image to the child “chef”. However, the seed image is not a picture of food but rather a picture of another unrelated subject, such as a muddy tree log lying on the ground with a line of ants walking over it.

The humanoid robot will say to the child, “Make me some food that looks like this!” and then begin observing how the child manipulates the objects and composes them. It will then be up to the child to use her available food objects to create the concept conveyed by the picture. We anticipate and will test in future experiments, that the child will create a new metaphor for the subject using the food items. The resulting metaphors may consist of a stick of celery (, or “log”) with soy butter (, or “dirt”) to glue the raisins on (, or “bugs”). The child may name the new food item, “bugs on a log”. The robot can respond with “Yummy! What do you call it?” and then begin the verbal exchange process to incorporate the novel concept along with the operators to produce the concept for future re-use.

3 Creative Concept Invention and Representation

We target the problem of metaphor learning in the context of a pretense play setting. In this setting, the robot presents a picture not involving food items to a child and asks the child to create a dish based on the picture. The underlying idea is that the child will see the objects in the picture as metaphors for the food items it has been given, and create a dish based on the metaphor. The robot then uses a description of the dish given by the child to recreate those metaphors itself.

A humanoid robot’s ontology consists of the concepts known by the robot, or agent, and the interrelationships between those concepts. The concepts can include attributes, actions, metaphors, and analogies. In a system in which more than one agent exists, or a multi-agent system (MAS), each agent, , has a set of concepts, , that it knows about.

\[
A = \{ A_1, A_2, ..., A_i \}
\]

Eq. 1) For each , there exists \( \Phi_i = K(A_i, \Phi_i) \)

A concept can be related to another concept, can define a class of related sub-concepts, describe a function on a set of objects, or describe a particular instance of a concept [Gruber, 1993]. We have chosen to use PyCLIPS, a python interface to CLIPS, to represent concepts known by the humanoid robot and to be able to reason about the concepts in the healthy snack food design game. CLIPS is a rule-based knowledge-based programming language used to represent and reason about knowledge [Giarratano and Riley, 1998]. It contains the three main components of an knowledge-based reasoning system: a fact list, knowledge base, and inference engine.

In the previous section we described the scenario in which the humanoid robot seeds the interaction with a picture of ants on a log. It will represent the concepts it knows as facts in its knowledge base. For a given relation name, our reasoning has to be told which are valid symbolic fields, or slots [Giarratano and Riley, 1998]. Groups of facts that have the same relation name and contain common information can be depicted using the deftemplate construct.

The particular ed image we are using in our example can be represented by the CLIPS facts:
knowledge base in one of three ways:

- (Eq. 3)
  ( defeacts picture-concepts
    (robot-concept (name nothing))
    (robot-concept (name ants))
    (robot-concept (name log))
    (robot-concept (name dirt))
    (robot-concept (name clouds))
    (robot-concept (name sky)))

The location of the robots concepts in relation to each other were represented using the following relations:

- (Eq. 4)
  ( defefacts picture-position-info
    (picture-position-relation (top nothing) (bottom ant))
    (picture-position-relation (top ant) (bottom dirt))
    (picture-position-relation (top dirt) (bottom log)))

The child will use the seed image and along with its set of healthy snack food elements it will attempt to create the image using only the list of available ingredients as described in the previous section. Although the child uses her own mental model and does not use CLIPS, the representation of the healthy snack food item that she creates is asserted as follows:

- (Eq. 5)
  (assert (human-concept (name nothing)))
  (assert (human-concept (name raisins)))
  (assert (human-concept (name celery)))
  (assert (human-concept (name soy_butter)))

These assertions come through a verbal interaction between the robot and the child in which the robot queries the child about the dish she has created. The locations of the objects used by the child in relation to each other were represented by the following assertions into the robot’s knowledge base:

- (Eq. 6)
  (assert (dish-position-relation (top nothing) (bottom raisins)))
  (assert (dish-position-relation (top raisins) (bottom soy_butter)))
  (assert (dish-position-relation (top soy_butter) (bottom celery)))

These facts would be incorporated to the robot’s knowledge base in one of three ways:

(a) by observing the child combine and stack the ingredients to make the food item
(b) by having the robot interactively inquire what the new food item has and how they are stacked, or placed, on each other
(c) using a combination of (a) and (b).

Our approach will make use of the Nao humanoid robot’s speech recognition abilities to conduct dialogue with the child to discover the new facts related to the food item using approach (b) [Gouaillier, 2008].

In order for the humanoid robot to discover these new metaphor concepts (e.g. raisin is a metaphor for ant), the CLIPS inference engine will be used to “fire”, or match, the appropriate production rules. For example, the rules to find the a new metaphor will have the following forms:

- (Eq. 7)
  ( defrule find-metaphor "Find a metaphor"
    (?f1 <- (robot-concept (name ?X))
     ?f2 <- (robot-concept (name ?Y))
     ?f3 <- (human-concept (name ?A))
     ?f4 <- (human-concept (name ?B))
     ?f5 <- (picture-position-relation (top ?X) (bottom ?Y))
     ?f6 <- (dish-position-relation (top ?A) (bottom ?B))
     =>
     (retract ?f1)
     (retract ?f3)
     (assert (metaphor (robot-concept ?X) (human-concept ?A)))
     (assert (metaphor (robot-concept ?Y) (human-concept ?B))))

The production rule in Eq. 7 was used with our reasoning engine to discover a new metaphor relation through the interaction between the robot and the child. This example rule should infer that the ant is stacked on top of the log as well as the dirt and that raisins are stacked on top of the celery as well as the soy_butter. This asserts these new facts. This rule will pattern match the contents of two concepts provided by the humanoid robot and the corresponding child concepts and discover the metaphors that the child creates so that they can be incorporated in the robot’s knowledge base. This production rule metaphorically should the concepts in the ant log picture provided by the humanoid with the concepts in the ant log food object. This rule was combined with other facts and relations to demonstrate the feasibility to represent and infer new metaphor concepts through the child-robot interaction in the health snack food design game. The results are discussed in the next section.

4 Evaluation

We executed our knowledge-based program using the CLIPS inference engine on a MacBook Pro with a 2.6 GHz Intel Core i7 processor and 16 GB of RAM. CLIPS is a rule-based reasoning engine that utilizes forward chaining to infer new knowledge.

The initial facts when the program constructs were loaded and the CLIPS system was reset included the following:

- f-3 (robot-concept (name log))
- f-4 (robot-concept (name dirt))
- f-7 (picture-position-relation (top nothing) (bottom ant))
When executing the program, the find-metaphor rule was fired with facts: f-4, f-3, f-13, f-12, f-9, and f-16. The resulting knowledge base contains the following concepts and relations:

- f-5: (robot-concept (name clouds))
- f-6: (robot-concept (name sky))
- f-7: (picture-position-relation (top nothing) (bottom ant))
- f-8: (picture-position-relation (top ant) (bottom dirt))
- f-9: (human-concept (name nothing))
- f-10: (human-concept (name raisins))
- f-11: (human-concept (name soy_butter))
- f-12: (dish-position-relation (top nothing) (bottom raisins))
- f-13: (dish-position-relation (top raisins) (bottom soy_butter))
- f-14: (dish-position-relation (top nothing) (bottom raisins))
- f-15: (dish-position-relation (top raisins) (bottom soy_butter))
- f-16: (dish-position-relation (top soy_butter) (bottom celery))
- f-17: (metaphor (robot-concept dirt) (human-concept soy_butter))
- f-18: (metaphor (robot-concept log) (human-concept celery))

The creative concept invention knowledge-based program discovered two new metaphors. The child’s concept of “dirt” was discovered to be a metaphor of “soy_butter”. Also, the child’s concept of “celery” was discovered by the robot to be a metaphor of “log”. These results are discussed in the following section.

5 Discussion of Results

Although there were two metaphors discovered and incorporated into the robot’s knowledge base, there was a key metaphor missing. We expected the robot to discover that “raisins” are a metaphor for “ants”. However, our knowledge base after the execution of our program failed to discover this key metaphor relation. In other runs of the program, incorrect metaphors were found but the resulting key find-metaphor rule was able to accurately uncover two out of the three possible correct metaphors. This error occurred due to the fact that the find-metaphor rule retracted fact assertions necessary for the inference engine to discover the additional metaphor for ants. Additional rules for finding metaphors will have to rely on more than location attributes and also be able to determine when nested relationships occur such as the one in our example.

Although our results demonstrate successfully that it is feasible to use our approach to invent new concepts through our metaphor-guided child-robot interaction in a health snack food design game, we see that improvement is needed. We believe that since our current program is based only on location relations, our system’s ability to find new concepts, or metaphors, can be improved by incorporating other key attributes such as shape or maybe texture.

6 Conclusions and Future Work

We have described a method for representing concepts in a metaphor-guided pretense play interaction and to infer new metaphor concepts using a reasoning engine and have demonstrated the feasibility of our approach. This paper describes a rule-based method to learning new concepts from observation. In this scenario, the robot initiates the learning by asking a child to ‘prepare’ a healthy snack based on a picture. The child has to make an interpretation of the image and come up with a novel concept. By formally annotating the original object with the novel concept and annotating the ‘prepared’ snack in the same way, the robot is able to figure out the metaphor and infer a new concept.

This paper describes a simplified and limited scenario which allows this knowledge-based system to computationally deal with the task. We have updated an earlier description of our general approach and test setting by adding implementation details for our concept representation and learning along with some preliminary empirical results. Further research is required to determine how far this approach will scale up with regard to the knowledge acquisition of new metaphor concepts. We plan to incorporate this approach into our prototype humanoid health coach as one of the activities to teach children about how to make and eat health food snacks. Our approach holds promise for humanoid robots that can interactively learn new metaphor concepts by interacting with children in food design scenarios.

Acknowledgments

This paper extends an earlier version of this concept from the AAAI 2013 Spring Symposium with further implementation details and preliminary empirical results for knowledge representation and reasoning to arrive at its current form. The authors wish to acknowledge the helpful comments from the reviewers that were incorporated into this paper.

References


Back-And-Forth Inception: Towards a Cognitively-Inspired Model of Concept Construction via Conceptual Intellection

Ahmed M. H. Abdel-Fattah and Stefan Schneider
Institute of Cognitive Science, University of Osnabrück, Albrechtsstr. 28, Osnabrück, Germany,
{ahabdelfatta, stefschn}@uni-osnabrueck.de

Abstract

Based on experimental evidence, we introduce our ideas of ‘inception’ and ‘intellection’, then we use the ideas to propose a preliminary model of concept construction via ‘conceptual intellection’. First, the experimental evidence is introduced and analyzed. Based on this analysis, we conceptually discuss how the construction of concepts can function in cognitively-inspired models, in which knowledge can be represented as concept entities.

1 Overview and Motivation

It has always been a big challenge for the interested researchers, in a spectrum that ranges from closely related domains to orthogonal and interdisciplinary ones, to give sentientious, exhaustive definitions of ‘creativity’. During only the past few years, a multitude of relatively new (interdisciplinary) theories and ideas on creativity are introduced, which add to the numerous psychological studies that have already been given during the past few decades. We refer the interested reader to the list of references given in [McCormack and d’Inverno, 2012; Boden, 1996, p. 267].

Recent theories on creativity excessively ground theoretical ideas on creativity through practical applications and practice, and describe many ways to understand creativity through computers (e.g. in the interdisciplinary volume [McCormack and d’Inverno, 2012]). Although this theoretical grounding is multi- and interdisciplinary in nature (since it bridges notions from computer science, psychology, neuroscience, visual art, music, and philosophy), such theories and studies still do not fully capture deep insights that actually help in building computational models of creativity (with a cognitively-inspired flavor). For one thing, the recent theories illustrate how computers are fundamentally changing what one can imagine and create. Unfortunately, the theories neither widely nor deeply focus on how human participants can affect the proposed computational models. This disregards that participants are very likely to have different knowledge, which affect what the participants imagine or create. For another, defining creativity and aspects thereof are extremely difficult, if at all possible, to express in a formal system, but can be described or explained in terms of the observed outcomes and related phenomena.

The interest in studying and developing computational models of human-like creativity renders it inevitable to take these observations into consideration when proposing similar theories. One of the main motives is to contribute to the latter need. We focus on how experimental results may influence the design of computational models, based on stimulating the way human participants construct their own understandings of conceptions. We specifically suggest a ‘life-cycle’ of such construction, which has the potential to be simulated in computational models, using computationally-plausible cognitive processes. Later, we propose a preliminary model that is based on this suggestion, but an overview summarizing the central ideas is given in the following.

Although humans are widely considered the best known exemplar of cognitive beings, they are born knowledgeless, then they acquire their knowledge from various sources over time. In principle, their knowledge is built of disparate pieces of information (e.g. beliefs or facts), which we call concept entities. These disparate concept entities can have implicit, rather interrelated, tying links, that enable humans to utilize their cognitive abilities in arranging the entities into conceptions; a network of conceptions that forms the background knowledge. Depending on different situations and contexts (i.e. experience), the existing tying links between concept entities can either be weakened, or strengthened. New links can also be built. As people contemplate these links, the conceptions themselves can be updated, and others may be created. Conceptions can be learned and updated either by such contemplation and deeper understanding, or by acquiring more knowledge. Both ways affect the already-existing knowledge conceptions, which results in the network of knowledge conceptions being re-organized. This goes on and on, which means that conceptions (as well as their composing entities) are being ‘intellectually’ treated and dealt with all the time: by updating already-existing conceptions, by creating new ones, or by re-arranging the hierarchy that possibly categorizes them and their interrelated entities. We will call this mechanism, of continuous intellectual treatment of the conceptual entities, the “intellection” mechanism.

We claim that the above view has the potential to be paralleled in artificial models. We will propose a model that mainly operates by employing a process that corresponds to the intellection mechanism mentioned above. In the model, we call this main process “conceptual intellection”, since it
stimulates, arranges, constructs, and updates concept entities and concepts. In other words, ‘conceptual intellection’ is the name we give to the main process (in the proposed model), which utilizes other processes in producing and updating concept entities and concepts. The article presents our views on how cognitive mechanisms are involved in ‘intellection’, and how the proposed model potentially simulates them. Note that some terms may have different uses according to different contexts. For example, we usually prefer to use terms like ‘mechanisms’, and ‘conceptions’, whereas, from the modeling point of view, ‘processes’, and ‘concepts’ (or concept entities) are preferred. In general, the mapping between the notions and the contexts is not one-to-one, though the meaning should be understood (e.g. a cognitive mechanism and its corresponding computationally-plausible cognitive process may both be termed ‘cognitive mechanisms’). ‘Conception’ is treated in most of the cases as the ideation in the human mind, while ‘concept’ is used when we are more interested in the representation issue. As already mentioned, we use ‘intellection’ for the human cognitive mechanism, and ‘conceptual intellection’ for the corresponding process in the model.

The rest of the article is organized as follows. We start in section 2 by giving evidence from cognitive science experiments that help us in backing up the above views and claims, by arguing how humans use their cognitive capacities in stimulating, creating, and updating their conceptions. In section 3, we state and explain our proposal of how the given evidence may help cognitively-inspired models to achieve concept construction: (1) we suggest general assumptions about this kind of models (cf. section 3.1), then (2) explain a ‘life-cycle’ of concept construction in details (cf. section 3.2). In section 3.3 we give quick hints on some related literature. In section 4, some concluding remarks are given.

2 Constructing Conceptions in Humans

Two case studies are investigated in [Schneider et al., 2013a; 2013b; Schneider, 2012] to show how participants construe conceptions, and possibly solve problems related to inductive production.

2.1 Stimulating Structure Construction: Study 1 (Algorithmic Structures)

A study in [Schneider et al., 2013a; Schneider, 2012] investigates how participants make up the regularities of a positional number system by way of example. A task in the study ‘disguises’ the surface features of the familiar number notation by using a quaternary system with the four symbols A, B, C, and D. This task is presented to human participants. The participants were only given the initial symbol sequence A, B, C, D, BA, BB, written on a tablet screen, then were asked to continue it (cf. figure 1). No further restrictions were put onto the participants, so that they can continue the sequence while creating their own conceptualization.

Two interrelated aspects of the findings are highlighted in the following. First, participants assemble Occurrence representations of the system as a ‘hierarchized’ network of metaphors of differing scopes, and, secondly, the temporary hierarchies scaffold the construction and change the participants’ conceptualizations. (N.B. Readers who are interested in other aspects of the study, such as the number of participants and the typicality of the presented data, need to refer to [Schneider, 2012] for more information about the study and the investigated group of participants.)

Given the sequence of letter combinations A, B, C, D, BA, BB (see figure 1-(a)), all the participants continued the sequence by appending BC, BD. The protocols in [Schneider, 2012] show that the task was apprehended as to create a ‘linear continuation’ of ‘two-letter combinations’. The way in which combinations were accomplished is by repeating the ‘ABCD-series’ at the right position while after each turn ‘prepending’ another symbol at the left. Some participants were irritated that the first prepended symbol was a B, not an A, and perceived this as an inconsistency, so their continuation showed four major types of coping with what they conceived as a challenge.

In figure 1-(b), three types are shown: (i) the continuation to the left regularly repeats the missing of a symbol; (ii) the branch in the middle defers to cope with the problem, which resulted in later complications (cf. figure 1-(d)); and (iii) the continuation to the right avoids the structural problem by appending an analogue version using the next letters of the alphabet. In all three cases, the hierarchical understanding is the basis of devising and realizing further continuations. The fourth continuation type is shown in figure 1-(c), where the aspect of expanding positions was put in focus. The main idea was that, after each repetition of the ABCD series, another symbol is being prepended. This leads to an expansion that one participant calls a ‘triangle’, another a ‘pyramid’, and yet a third associates the related metaphor of ‘atom splitting’.

Drawings and gestures (which are being depicted by arrows in figure 1-(c)) also show this figurative system comprehension. The pyramid metaphor only gives a coarse plan of what to do when positions get expanded, but as such, it sets the frame for ‘filling in’ detail and ‘extracting’ other aspects.

We now need to summarize how construction guides the participants’ coping with the task, since this will turn out to be highly related to the proposal given in the next section (cf. section 3).

Participants naturally construct an understanding consisting of structures of differing scopes, when they apprehend the given symbol sequence and the underlying generative principle. In general terms, one may speak of the ability to construct and develop a conception by the continuous grasping and organization of both (i) the entities that constitute a given stimulation about this conception, and (ii) the underlying connections between such entities. Moreover, more comprehensive structures guide their instantiation. As a result, the whole construction determines its further application and development. Structures tend to subordinate and re-arrange other structures, and the degree of comprehensiveness does not restrict which structure is being modified. If an instantiation (inception) of one’s conception is not readily at hand, modification of the assembled components may either change the detail structure or accommodate the comprehensive structure. For example, the case in which a ‘pyramid’ of letters is expanded indicates that, when one focuses on a stimulating
aspect (the inception of position expansion) one may bring up a conception (e.g., the pyramid metaphor) and thereby modify an earlier one (e.g., the linear succession). The new conception then guides in constructing further detail (e.g., filling center digits). Similarly, figure 1-(d) indicates that the deferral of the A-problem leads to completing the linear sequence with AA, AB, AC, AD, followed in turn by BA, BB, BC, BD. This modifies the linear continuation conception to become circular, or loop.

2.2 Stimulating Structural Construction: Study 2 (Mental Imagery Construction)

Imagery is widely considered to be essential to creative processes [Finke et al., 1989]. The evidence given in [Schneider et al., 2013b] show that imagery does not operate with static images, but requires to frequently reconstruct them. Apparently, one can mentally assemble and recognize a figure; a process which is, according to Hofstadter, based on “higher-level perception” and thought of as “the core of cognition” [Hofstadter, 2001]. The assembly suggested in figure 2 is given by verbal description in [Schneider et al., 2013b], where the task is to count all triangles in it. Participants easily found some of the triangles, but struggled to find all. They readily ‘remembered’ the whole figure; but apparently had to ‘reassemble’ varying parts of it in order to consider ‘relations’ between lines that would ‘form’ a triangle. In the presented task, assembling parts is guided in that a prototype – a triangle in this case – ‘consummates’ relevant material. In the case where a prototype dominates thought development we want to speak of construction as a special type of assembly. When such a construction fails (and nothing is being recognized), one attempts a new construction.

2.3 Aspects of Construction via Intellection

The first study investigated the inductive learning of a concept. In study 2, the participants’ reports indicate that mental images are situations that need thoughtful intellection. That is, intellections that have to be construed, and continuously reconstructed, to discover relations. In both studies, participants construed and operated with an intellection of the given stimuli representing different aspects of it. Thus, an ‘intellection’ here means a structurally hierarchized apprehension of a situation, which gives an immediate directive of what could be done. An intellection is being carried out until the ‘step’ where it fails to ‘proceed forward’ (in the model, this corresponds to a concept being produced). One then has to reconstruct and test the new ideation, where reconstruction usually starts with a very similar approach like before. The process can be viewed as an attempt to ‘fulfill’ the specification given by the initial stimuli. The construction does not search and find, but stimulates and instantiates a certain specification (inception) and thereby creates something new. Mutual modification of structures of different coarseness is another, essential property of intellection construction. Construction is guided by comprehensive structures, but a comprehensive structure can also be modified as a result of working on structure details.

In the model, proposed below, we view ‘conceptual intellection’ as an open-ended process that is intended to computationally simulate the ‘intellection’ mechanism discussed above: the latter is our view of the way in which humans construct and update their conceptions (based on the analysis of the given experiments and the motivational view in section 1), and the former is our view of the way the proposed model simulates the latter. The construction and update is mainly performed by means of the existing knowledge concepts and new stimulating entities. These stimulating entities cause the process of inception to take place, again, and start a conceptual intellection process again. This back-and-forth processing is thoroughly discussed in the following section.
3 Cognition-Based Concept Construction

From a cognitive science point of view, we claim that creative production of conceptions, in the human mind, is a (potentially unbounded) mechanism of cognition that utilizes simpler sub-mechanisms in the construction and organization of conceptual entities. So the simulation of this mechanism (in a computational model of creativity, for instance) would be a process that needs to reflect the functionalities of other interrelated (cognitive) sub-processes; each of which is

1. intermediate: an ultimate (creative) production has to go through a non-empty set of the sub-processes, and
2. interacting: one sub-process, that may be involved in a construction process, affects the functioning and outcome of another.

In what follows, the previously discussed aspects are transformed into a proposal of a model that can eventually accomplish (creative) production of concepts. Based on the way conceptions are constructed in humans (the experimental results and our claims in section 2), the model views (creative) concept construction as a life-cycle that utilizes cognitive processes in carrying out conceptual intellencelations.

3.1 General Model Assumptions

We do not want to restrict our presentation to one specific model or another, because the conceptual ideas here are intended to apply to a range of models that satisfy some assumptions. We will not be designing a computational model of creativity, rather proposing our view about how cognitively-inspired, computational models may employ cognitive processes to achieve (creative) production of concepts. Building this kind of models needs to reduce computing (creative) construction of concepts, in particular, to computationally-plausible cognitive mechanisms. For example, an agent (of a model that satisfies the assumptions) can update an already-existing concept, say $C_T$, by making an analogy to a well-known concept, say $C_S$. The update can happen because an analogy-making may involve an analogical transfer that enriches the ‘structure’ of $C_T$ (humans would also update their understanding of $C_T$ in a similar analogy-making situation). Another example is the utilization of concept blending in creating new concepts and recognizing unprecedented concept combinations.\footnote{Although these examples and figure 3 exemplify some suggested mechanisms, such as analogy-making and concept blending, we do not intend here to give a list of the cognitive mechanisms. Related discussions can be found in [Abdel-Fattah and Krummack, 2013; Abdel-Fattah et al., 2012; Abdel-Fattah, 2012].}

General assumptions about how we conceive the model are given in the following.

Let us first assume that one’s ultimate goal is to design a cognitively-inspired model of computational creativity and general intelligence. We also assume that the considered model provides the ability to represent knowledge in the form of conceptual entities. The organization of knowledge in some form of concept domains follows the language-of-thought hypothesis [Fodor, 1983], where concepts can be provided in modular groups to serve simulating several processes. A model of this kind helps the acquisition, organization, and development of knowledge using concept entities.

This aims at imitating aspects of knowledge emergence in humans that are based on their ability of acquiring knowledge from the environment, and using their experience in refining their understanding of repeatedly-encountered conceptions (cf. [Abdel-Fattah and Krummack, 2013; Abdel-Fattah, 2012]). In any case, we assume that the storage and organization of knowledge in structured concepts is unavoidable for showing human-like thinking or creative behavior by this model type.

3.2 The Life-Cycle of Conceptual Intellencelations

A proposed life-cycle for concept construction in artificial systems is explained below and illustrated in figure 3.

According to our proposal, a conceptual intellection computationally simulates a (creative) construction of conceptions; a (fulfillment) mechanism that is always incremental and potentially unbounded. The system should always build and modify its (created) results, be they concept entities, concepts, or arrangements thereof. The act of grasping conceptions caused by the stimulating entities (in a lively active human being’s mind) may never stop. So, a point where (creative) construction stops should be generally undesirable in a computationally running system. This does not have to conflict with notions concerning boundedness (e.g. bounded rationality [Gigerenzer and Selten, 2002; Simon, 1984], knowledge and resources [Wang, 2011], etc.) because these notions are concerned with how the cognitive (natural or artificial) agent thinks and behaves as the limited resources only allow (e.g. the processing and storage capacities).

Based on the given studies, we propose that the life-cycle of a conceptual intellection (that is, of constructing and updating the concepts) consists of four facets: (1) inception, (2) arrangement, (3) production, and (4) adaptation. Each facet constitutes specific cognitive processes. The depiction in figure 3 metaphorically uses a staircase object\footnote{In fact, the staircase object, depicted in figure 3, ought to be seen as a spiral of repeating stair-steps going up indefinitely. This is also illustrated by the spiral and the up-arrow, drawn to the left of the staircase object in the same figure.} with four sides; each corresponding to a facet of the potentially unbounded, construction process. Each facet’s stair-steps metaphorically correspond to interrelated cognitive processes (not all of them can be shown), which may be involved in this facet’s functioning.

The utilized metaphor in the given figure tries to pictorially deliver our own message that:

The more a system uses knowledge during the progressive inception (from brainstorming to identification of context-dependent pieces of knowledge), the more the system organizes and arranges the collected knowledge (during the progressive arrangement) before the ultimate (creative) productions are constructed, then finally adapted to the overall cognitive state (before a new life-cycle starts).

Using stair-steps implicitly reflects that the ‘proceeding’ through the staircase may go both ways at some times, causing some processes to be retracted and others re-executed.
At the beginning of a conceptual intellection life-cycle, stimulating entities are instantiated. The stimulating entities may have already existed (then activated somehow), or received as new input. They can have structural relations tying them together (or to other knowledge entities). The inception facet initiates the construction of concepts, which will lead to the creation of a new concept, or the modification and the hierarchical arrangement of already-existing concepts. We elaborate more on the cycle’s facets in the following:

1. The ‘inception’ facet involves activating the set of processes that leads to implanting and updating conceptual entities (which form the agent’s knowledge concepts; cf. section 3.1 and [Abdel-Fattah and Krumnack, 2013; Abdel-Fattah, 2012]). The inception can eventually result in the model having its set of concepts updated. This is because either (i) the existing conceptual entities are modified, or (ii) new entities are acquired. In the former case, the links tying some of the conceptual entities are affected and, consequently, the concepts involving them are also affected. The concepts can also be affected in the latter case, and new concepts may also be created.

2. The ‘arrangement’ facet follows the ‘inception’ facet, and is responsible for organizing (currently existing) conceptual entities. (Refer to the discussions in section 2.2 about hierarchical assembly and the attempts to re-construct when the assembly fails.)

3. The ‘production’ facet starts with the ‘construction’ process, which constructs a concept (after assembling and arranging its components) and ends with such concept having its components enriched (the fulfillment of the initial model specifications; cf. section 2.3). As a process, ‘construction’ is not the crowning result of the ‘production’ facet, rather the starting step (in a given cycle). Newly constructed concepts may be re-constructed and enriched in several ways (e.g. through analogical transfer). It is worth noting that, in the current proposal, ‘what counts as a creative construction and what counts as not’ is not specified and left to another discussion. This is mainly because the knowledge concepts can always be updated, though not considered new or ‘really creative’ concepts.

4. The fourth facet, ‘adaptation’ parallels the deeper, insightful understanding of conceptions (by humans), and reflects the changes in an input or a current context. There is always the possibility that what might have once been considered just a recently produced concept by the model (e.g. as a result of conceptual blending), becomes a well-entrenched concept later (perhaps after repeated usage).

Digression – The Never-Ending Staircase

The visualization in figure 3 makes use of an attention-exciting construction instead of, for instance, drawing a dull circle in two dimensions (which might have as well achieved the same representational goal). It may be interesting for some readers to see classical ideas represented in unusual ways, but there is more to this issue than only drawing an unusual, seemingly-complex shape.

A shape of an impossible triangle, famously known as the Penrose triangle, was originally created for fun by Roger Penrose and Oscar Reutersvärd. Later, Roger Penrose and his father, Lionel Penrose, created the impossible staircase, and presented an article about the two impossible shapes in the British Journal of Philosophy in 1958 (see [Penrose and Penrose, 1958]). M. C. Escher artistically implemented the two shapes in creating two of his most popular (and very creative) drawings; namely: “Waterfall” (lithograph, 1961 [Hofstadter, 1979]).

The idea of entrenchment has been elaborated on in [Abdel-Fattah and Krumnack, 2013; Abdel-Fattah, 2012].
3.3 Related Literature

In addition to the earlier discussions, the ideas of the proposed model are further supported by work in the literature. The suggested (four-facet) cycle agrees with the “Geneplor model” [Finke et al., 1992], which proposes that creative ideas can only be produced in an “exploratory phase”, after an individual already constructs mental representations, called “pre-inventive structures”, in the first “generative phase”. Such pre-inventing, in our view, initiates the inception facet and finishes before creative construction starts. Usage of the term “inception” in our proposal is thus akin to “pre-inventing”. Additionally, evidence is given in [Smith et al., 1995, p. 157–178] that the already-existing concepts greatly affect the (re-)structuring of the newly developed ideas. Moreover, Weisberg argues that (human) creativity only involves “ordinary cognitive processes” that yield “extraordinary results” [Weisberg, 1993], which consolidates our motivation to focus on cognitively-inspired models.

On the one hand, there are similarities between our proposal and the mental models theory, as proposed by Johnson-Laird and others (cf. [Johnson-Laird, 2006; 1995; 1983; Gentner and Stevens, 1983]). The reader may for example recognize that our view of what we call “intellections” seems analogous to Johnson-Laird’s view of what he calls “models”. “Models are the natural way in which the human mind constructs reality, conceives alternatives to it, and searches out the consequences of assumptions” [Johnson-Laird, 1995, pp. 999]. In addition, the general results of the many psychological experiments given in [Johnson-Laird and Byrne, 1991], for example, may be considered yet a further support to our proposal about how humans construct their understandings. On the other hand, our approach differs from the theory of Johnson-Laird in at least the following essential points. Whilst mental model experiments are mainly concerned with testing one particular ability of participants (namely, their ability to reason ‘ductively’), the first study in section 2, for example, investigated the ‘inductive’ learning of a concept. In our view, asking the participants to construct arguments, to show that one conclusion or another follows from what already lies within what is already given, is different from giving them inceptions, then allow them to freely continue their conception constructions. Creativity is revealed in the latter case, and not necessarily in the former. In fact, it is not even desirable that the participants reflect a creative ability (according to the aims of the tests in the former case). Another difference is that our proposed model has the potential to utilize computationally-plausible cognitive mechanisms in achieving artificial systems that realize the model. We are neither concerned with only testing deductive reasoning, nor do we need results that only obey rules of deductive reasoning. Rather, our main concern is in testing (creative) concept construction, which seems to involve cognitive capacities other than deduction. For instance, concept blending possibly allows for unfamiliar, resulting concepts by blending two (or more) familiar concepts, opening thus the door to non-deductively achieving thinking outside the box. Also, given two concepts, people do not usually see one, and only one, analogical mapping between the two concepts’ entities. Counterfactual reasoning is yet a third example that emphasizes the importance of utilizing mechanisms other than deductive reasoning in (creative) concept construction (actually, deduction may not at all work when computationally analyzing counterfactual conditionals [Abdel-Fattah et al., 2013]).

4 Concluding Remarks

It is not surprising that many ideas, of what currently is regarded standard in (good old-fashioned) artificial intelligence (AI), have already been laid out in Newell and Simon’s article [Newell and Simon, 1976]. The article very early anticipated improvements that seem inevitable from today’s perspective. According to Newell and Simon’s pragmatic definition, an intelligent organism is required to be ‘adaptive’ to real world ‘situations’. Newell and Simon inspired the scientific society to begin studying human-like intelligence and their ability to adapt. However, since the beginning of AI, it was “search” what they proposed as a cornerstone of computing [Newell and Simon, 1976]. Less importance has been given to other computationally-plausible cognitive mechanisms (which may not need, or be based on, search).

Creativity, on the other hand, is a complex general intelligence ability, that is closely connected with the utilization of various cognitive mechanisms. Creativity may definitely need searching in performing specific sub-processes, such as finding analogies to newly available concepts before modifying (or adapting) them. But, after all, it certainly differs from mere search in several aspects. Our trial in this article claims that investigating, testing, and implementing cognitive mechanisms need to attract more attention, in particular when one is interested in modeling computational creativity.

A model of computational creativity is expected to greatly differ from a classical AI model. An essential difference between creativity and AI is that AI programs have to access huge amounts of (remarkably meaningless or unsorted) data before getting (meaningful and interconnected) results, whereas creativity reflects the general intelligence characteristic of humans, who are much more selective and adaptive in this regard. For example, humans can not only induce appropriate meanings of unknown words, both within and without a given context (e.g. interpreting unprecedentedly-seen noun combinations that appear in general phrases; cf. [Abdel-
tions or modify them, using the already-existing knowledge of an (artificial) intelligent system to construct computational creativity should be phrased in terms of previously known constructs to new contexts. We propose that potentially unbounded construction process, without crossing the exponential bounds that search may require. Such a potentially unbounded process of construction is proposed here as a ‘looping’ process. It is therefore worth quoting part of Hofstadter’s comments on what he calls “the central cognitive loop” (described in [Hofstadter, 2001]):

“The broad-stroke pathway meandering through the limitless space of potential ideas [...] goes around and around in [...] a loop, alternating between fishing in long-term memory and unpacking and reperceiving in short-term memory, rolls the process of cognition.” [Hofstadter, 2001, p. 521]

From where the authors stand, creativity produces uncommon outcomes that provide contemporary applications of previously known constructs to new contexts. We propose that (computational) creativity should be phrased in terms of the ability of an (artificial) intelligent system to construct conceptions or modify them, using the already-existing knowledge about former concepts; a process that is both computationally plausible and cognitively inspired (cf. [Abdel-Fattah and Krumnack, 2013; Abdel-Fattah et al., 2012]), though potentially unbounded. Such systems treat the behavior ‘to create’ as that of bringing into existence a concept that has not previously ‘experienced’, or that is significantly ‘dissimilar’ to another analogue (or even its older version). But, in essence, creativity cannot be about creating something from nothing (computationally-speaking, at least), rather triggered by ‘inceptions’ every now and then.

Acknowledgments

The authors would like to thank the three anonymous reviewers of this article’s first draft, both for their valuable opinions and encouraging comments. The concerns they raised remarkably helped in improving the presentation of the given ideas in some parts of the article.

References


\*Again, the precise list of cognitive mechanisms can not be extensively discussed in this article.


Square, zero, kitchen: Start
Insights on cross-linguistic conceptual encoding

Francesca Quattri
The Hong Kong Polytechnic University
Hong Kong
francesca.quattri@connect.polyu.hk

Abstract
In this study, I approach the study of metaphor and shape from a linguistic/cognitive perspective. A distinction is made between metaphors of shape and metaphors of shaping. Two major case studies are presented (going back to square one [i.e. to start something anew because of a failure] and vicious/virtuous circle) with the shape in the metaphor’s source domain. ‘Metaphor’ is intended here as form of speech with collocational/idiomatic behavior; the source and target domain frame is experimentally applied to an idiom and not just a metaphor.

The study is qualitative, in progress and limited to two case studies, but it offers insights on the problem of conceptual encoding as observed from a new perspective (metaphors with shape words in the source domain) and values the literal and not just conceptual meaning of metaphors.

The final goal of the project is to help AI experts in their struggle to find robust models of representation of metaphorical thinking, conceptual encoding, as well as in-depth text and context analysis.

1 Introduction
As [Barnden, 2008] points out, the interest of AI to metaphor and similes dates back to several decades. Researchers including Russell 1976, 1985, Carbonell 1980, 1982, Norvig 1989 and Way 1991 have paved the way of investigation in the field.

The author of this paper aims at underscoring metaphors by the study of source-domain reasoning and backs up with citations from cognitive science, another aim of AI (in [Barnden, 2008]’s words).

Metaphorical thinking remains, despite outstanding research, a big black hole for cognitive and neuro-scientists as well as AI experts. [Ramachandran & Hubbard, 2003][Ramachandran & Hubbard, 2011] claim that the backbone of all metaphors is to be found in synesthesia (from the Greek roots syn, meaning “together,” and aisthesis, or “perception”), i.e. the ability to combine senses together given their spatial

1 Patients affected by chronic synesthesia cannot avoid to i.a. define numbers with colors and colors with personalities [Ramachandran, 2011] without making it up (e.g. “mint tasted like a cylindrical marble column”; “I enjoy music that has very metallic lines”; [Cytowic, 1993]:32f.).
Johnson, 1980][Koevecses, 2010] and plenty of studies on shapes as geometrical entities for automated detection (e.g. [Fonseca & Jorge, 2000]), or as a topic for discussion on formal description and ontology (e.g. [Borgo & Kutz, 2011]).

I present hereby two case scenarios. In the first, I analyze the metaphor (vicious/virtuous circle) and want to show that the perception of the shape greatly depends upon the fact of whether the metaphor is of shape or of shaping. In the second case (going back to square one), I try to apply the source-target domain parallelism to the idiom. Aware of the fact that this approach might be biased by some linguists, I define the case “experimental” and foresee to lead attention instead to at least two facts that have implications in conceptual encoding (and thus in the development in AI of sound models for the same).

The first observation is the ability of the human brain to process abstract information by means of some more abstract (and not by means of a simple paraphrase). When asked to declare the meaning of an idiomatic conventional form, other abstract forms are used.2

The second is the collection of observations on the literal meaning of the metaphors, regarded as relevant information carrier (sense versus meaning, description follows).

2 Metaphors of shape and metaphors of shaping

The perception of shape contributes to the understanding of metaphors. In particular, I claim that the visual perception of shape is strongly related to the fact of whether a metaphor is defined of shape or of shaping.

Notice that by ‘metaphor’ I define a figure of speech which might contain idiomatic/collocation meaning and form. Without going into details, the claim here is that the traditional linguistic division between idioms, collocations, metaphors, metonymies seems to be ad hoc and gives little indication of the true entanglement of these forms together.3

I define by metaphor of shape is a figure of speech where the entailed shape (in the source and/or in the target domain) has to be understood literally (as precise form with strong visualization details), in order to fully grasp the meaning of the expression itself.

Take for instance the sentence: “at a certain point in her life, Mary decided to draw the line”. Listeners cannot but imagine a rather clear line which marks a cut, a division between what it was and what is near to come. “I have not signed on the dotted lines yet, but will do so soon”: Also in this case, the expression (which means: formally agree on something by putting the own name on it) allows us to literally imagine a dotted line. These two examples are in my definition metaphors of shape.

Metaphors of shaping are, as readers might intuitively understand, figures of speech that give indication of the process of a shape coming into being, or, more importantly, the development of an event/object. The shape does not play a relevant role in the metaphor. It might be justifiable by etymology, but its origins (which explain why it was originally used for the form) are not always traceable.4 “Invest in lines, not in dots” is a common suggestion among market dealers. Obviously, the shapes are not used here as forms (‘lines’ are robust companies with history to count on; ‘dots’ stand for fresh start-ups with no experience on the market). Eventually, the study of shapes in metaphors can be extended to “compound” metaphors, like in the idiom kick the bucket. The metaphor is entailed in ‘bucket’ (which stands for coffin) and means to die, to pass out. Another example of compound metaphor is the expression “the circular triangle” or in the idiom “to square peg in a round hole”.

Despite having anticipated the concept of compound metaphors, I focus here on shape words as used in the source domain. It might be useful to notice that the conceptual pairing between source and target domain is for some scholars not always a sound way to define how metaphors work. [Agerri et al., 2007; Agerri, 2008] argues for instance that it is not always possible to establish a link or parallelism between source and target domain to explain a metaphor, since the expression is a highly contextualized linguistic phenomenon that derives much of its information from the surrounding discourse (“enrichment of mappings”. [Agerri, 2008]). The computational process deriving from these mappings can be endless and provides little help for a clear investigation of the subject.

[Hobbs, 1992] talks about metaphors in terms of relativity theory. The scholar experiments with a novel and a conventional metaphor. In the novel (“John is an elephant”), he tries to come up with a model to unambiguously map source and target domain and reasons on the properties that both John and an elephant may share (e.g. clumsiness). He does so by citing different approaches undertaken in the analysis of metaphors as proposed by i. a. Orthony et al. (1978), Glucksberg and Keysar (1990), Carbonell (1982), but claims that they are all a-contextual. Rather than (just) picking up the size or other physical properties of an elephant and compare them to John, Hobbs suggests that an analyst should consider the metaphor within the clause and within the situation of use “Mary is graceful, but John is an elephant” ([Hobbs, 1992]:15). In a way, Hobbs validates the Sapir-Whorf hypothesis of relativism, which claims that that there exists linguistic relativity in common sense language due to cultural contextual influences on language itself (a statement also indirectly sustained by [Pinker, 2009a][Pinker, 2009b][Pinker, 1994]). The same theory is defended by [Polley, 2010], who has proven in his doctoral dissertation the existence of emotion metaphors in English and Mandarin with varying degrees.

2The statement is not supported by experiment but by simple observation derived from a lexicographic search of the forms in different repositories. The viewer is naturally redirected to synonyms of the forms, which are also conventional idiomatic expressions.

3Further planned studies aim to find out whether these linguistic forms are not just parented with each other, but whether they are bound by a sort of hierarchy (from collocations derive metaphors [“the sharp bite is sat across the table”], some metonymies are metaphors [“he drank a glass of bitter memories”], from metaphors derive idioms (“she’s in a good shape”).

4A good archive for etymology of idiomatic form is the website The Phrase Finder, www.phrases.org.uk.
in the patterns of organization and use of these metaphors across languages and individuals. The study, focused on 幸福 (xīnghù, happiness) in Mandarin and English, is conducted through corpus analysis and adopts elements of the MPA, Metaphor Pattern Analysis by Stefanowitch (2006) and of the Metaphor Identification Procedure (MIPVU), as developed at the Vrije Universiteit by Stern et al. in 2010.

Ahrens et al. [Ahrens et al., 2003] defends the opposite front and proves, starting from Lakoff’s Conceptual Metaphor [Lakoff and Johnson, 1980], that all source domain can be put in correspondence with a target domain automatically and with little ambiguity. The authors integrate the Conceptual Mapping Model with ontology representation [SUMO, 2013] to defend their thesis. Several studies on ‘economy’ (jīngji) in Mandarin Chinese [Academia Sinica, 2006] prove that an automatic matching is feasible [Ahrens et al., 2003; Chung and Ahrens, 2006].

To make the Model work, a Mapping Principle Constraint is introduced, for which a target domain will select only source domains that involve unique mapping principles. In this way, ECONOMY in Mandarin Chinese is WAR, where WAR is subset of the source domain of COMPETITION. In another example, IDEA, a target domain, builds upon the source domain BUILDING, and not for the wiring, or plumbing system of a building, but for its being a stable block with pillows and corner stones. “Buildings involve a (physical) structure and ideas involve an (abstract) organization” [Ahrens et al., 2003].

The authors come to these conclusions via a three-step process: (a) first define entities, properties, and qualities that exist between source and target domain; (b) then go back to the source domain and compare it to the real-world knowledge; (c) eventually postulate a unique mapping, as the one presented above for building and idea. We decide to adopt the same process in our study for its clarity.

3 First case study: Vicious/virtuous circle

Given that the research is currently in process, only the major outcomes of the study on the metaphor are reported. The research has been conducted in different stages, with different granularity towards the syntactic elements surrounding the expression (verbs, adjectives, prepositions) and with regards to different domains of use.

The corpus-based analysis has been conducted on both [OCOA, 2013] and [SketchEngine, 2007]. For the latter, the corpus enTenTen has been selected given the high number of tokens provided.5

For each expression, I have worked on the sentences automatically filtered by [SketchEngine, 2007]. In previous analysis in fact, it has been noticed that going through the corpora manually (having set a number of sentences to analyze) resulted in having many examples to support a theory about the behavior of the metaphor and less examples for other cases, which could therefore be considered the exception. To avoid this and having to deal with as many applications of the metaphor in contexts as possible, I have proceeded with a random selection of examples. (around 1,700 entries for vicious circle and virtuous circle respectively (last search: April 21st, 2013).

The KWIC size has been set to 500 characters, which gives around three sentences before and after the sentence in which the collocation is entailed.

The study of the metaphor is run always with the intent to acknowledge whether the shape circle plays a role in the final understanding of the expression (in other words, whether vicious and virtuous circle are metaphors of shape or metaphors of shaping).

Given the collocation form of the metaphor, I expect that vicious and virtuous are adjectives referring to morality and virtue. I can suppose that since the circle is also a religious symbol, the choice of the adjective is theologically motivated. Crosslinguistically, if I compare the metaphor as it exists in other languages, I bump into the same conclusions: “Teufelskreis” (Ge, lit. “circle of the devil’); “círculo vicioso” (It; lit. “circle of the vice’; the same in the Portuguese and Spanish, “círculo vicioso’); “恶徳循環 (èxìngxūnhuán)” (Ch; lit. the evil circle).

Corpus-analysis applied on the form seems to revealed that the literal meaning of the shape has no influence on the final understanding of the metaphor. The expressions come often in contexts where different events are listed not in a random, but mostly in a cause-effect relation, or associated to verbs of rotation and motion. This suggests that the human brain does not need the shape of a circle to understand the metaphor, but it needs a close frame and a sense of motion to give it a meaning.

The actions involved with the circle are often: one or something get stuck in a circle; the circle can break, arise or simply be. The sequel of events follows a dynamics. In many cases this could be called causality, but there needs more evidence for this. As for the other question, whether the property of being vicious or virtuous affects the behavior of the metaphor in context, in none of the cases I have found that there exists an explicit reference between the shape with morality or immorality. Vicious seems to stand more for negative, virtuous for positive.

It seems that the metaphors of shape vicious/virtuous circle acknowledge the existence of a form in the source domain for two major reasons: an ontological function (the form stands for mappings, enclosure of facts, boundness, network), and an orientational function (the close frame suggests round directionality or causality more than a casual trajectory). The image of the circle is from absent to vague, definitely not significant for the metaphor.

Eventually, as stated above, it seems that the literal understanding of the adjectival form (morality, religion) is not relevant for its conceptual use in context.

If someone would ask to be explained the meaning of vicious circle, it is probable (at least by native language speakers) that the metaphor is explained via other metaphor or abstract forms.

Vicious circle has in fact the same meaning of “daily grind”, “treadmill”, “hamster cage”.

5The Ten Ten English web corpus, 3,268,798,627 tokens, 2,759,340,513 words.
‘Circle’, ‘treadmill’ and ‘cage’ have at least two aspects in common: (1) they are closed forms, especially ‘circle’, ‘grind’ and ‘cage’; (2) they rotate (‘grind’ is the product and the action of a treadmill).

4 Second case study: Going back to square one

The second expression considered is going back to square one. This form is an idiom.

Given the intertwined nature of linguistic forms (explained above), it also entails the collocation square one and the collocation back to or back at square one.

The attempt at this stage is to transfer the source-target domain logics of metaphors (which works fine for some scholars) to idioms, in the higher goal to automatize the process.

Ahrens [Ahrens et al., 2003] suggests a three-step approach for the identification of a metaphorical mapping. A quick recap of the stages: (a) define entities, properties, and qualities existing between source and target domain; (b) go back to the source domain and compare it to the real-world knowledge; (c) postulate a unique mapping, as the one presented above for building and idea.

An answer to the question: Why decide to adopt this model on idioms? The analysis of vicious/virtuous circle has showed recent patterns that cannot be totally proven until the research switches to statistical methods and more data. The outcome of the proposed research and of further researches on shapes already shows nevertheless that one single shape might happen in a metaphor of shape and in a metaphor of shapping, something being perceived as one or the other in one same form. The dizziness is justified by the context, which eventually determines the meaning of a metaphor per situation (without exception for conventional forms).

As in the current shape, the research reiterates what [Hobbs, 1992] and other authors already previously cited tell us about metaphors in context, namely the happening of the Sapir-Whorf theory of relativity.

Ahrens proves that an automatic source-domain detection for metaphors is feasible, under certain constrains. So my question is: Does it apply to idioms as well? Since I have adopted a definition of metaphor which enables the extension of the research on linguistic forms than metaphors, I test it on an idiom with shape.

The approach has been adopted with further variations: The whole idiom stands for the source domain (going back to square one); the target domain stands for the meaning/s associated to it. This enables to identify stage (a) of the method proposed.

The third passage of the experiment can be anticipated as well. The mapping between source and target domain is derived from encyclopedic knowledge: going back to square one means ‘start something anew because of a failure’.

What needs to be found are the qualities and relations between the two domains. The analysis can be done at several levels of granularity. The properties between source and target are either derived from the domains as we have defined them, or they are extracted from every single components that make up each domain.

Following the first option, little relation is found. If going back to square one means start something anew because of a failure, there are apparently no other linkings between the two sequences of information rather than the motion verb going back associated to the motion verb start/begin.

The second level of analysis provides more information, but it is unbalanced. We can in fact finely decompose the source domain, but not the target domain. Even in Levin’s verb classification [Levin, 1993] little information is given on the target.

The word square is polysemous. As a noun, it means inter alia “any artifact having a shape similar to a plane geometric figure with four equal sides and four right angles”, “a hand tool [...] used to construct or test right angles” or “something approximating the shape of a square”, [WordNet Search 3.1., 2010].

The collocation square one acts as collocation in certain domains of use, while in others it becomes a proper noun (a jargon term) with exact meaning (as in the case of football and rugby games, board games and the playground game hopscotch). This implicitly means that, even if we have decided to take it as an idiom, the expression going back to square one can be idiomatic or not according to the context. Evidently, the corpus-based analysis to be conducted will not take into account the form as a jargon.

Suppositions on the form might lead to think that going back to square one involves a return to a precise shape, precisely a square. The numbering (“one”) suggests that there may be other numbered forms apart from that. The action of going back involves a previous departure, ergo it may refer to a temporal frame. A spatial setting is also involved, with a movement from space A to space B and back to A in the simplest visualization.

The next question is: How does this fine-grain information relate to the concept or meaning of ‘start’? The verb entails a temporal and spatial location, but it would be a speculation to argue that a U-turn is also involved. That justifies the statement about an unbalanced conceptual mapping.

A final question relates to the choice of ‘square’ for building this idiom. Why is the concept of ‘start/begin’ precisely related to this geometric form (as it is ‘square’ taken in one of its many senses)? Does the concept need to relate to this form in particular? The attempt to find reasonable answers justifies another variation in Ahrens [Ahrens et al., 2003] study.

---

6 Back to square one: 1 bin 10 ml results, Google; back at square one: 1 bin 20 ml results, Google.
7 If you are back to square one, you have to start working on a plan from the beginning because your previous attempt failed completely’, [Cambridge, 2013]

8 The verb ‘start’ belongs to the category of the begin verbs, which describes “the initiation, termination, or continuation of an activity” ([Levin, 1993]:274-275). Since Levin’s work is not concerned with sentential complement-taking properties, no specification is provided about a possible distinction between instance ‘start’ and ‘begin’. ‘Failure’ as a verb is not mentioned.

9 From The Phrase Finder, http://www.phrases.org.uk/meanings/back\%20to\%20square\%20one.html
extend the research to a cross-linguistic comparison.

4.1 Cross-linguistic comparison

A random sample of 250 sentences is taken [SketchEngine, 2007] from the enTenTen corpus (Web corpus, 3,268,798,627 tokens, 2,759,340,513 words). The reason why random and not manually picked examples have been taken is the same as for vicious/virtuous circle.

In this first stage of the analysis I am interested to find out: (a) whether the idiom exists in other languages and (b) whether it is translated by means of the same form. In brackets the English literal translation. The expressions have been taken from [Linguee, 2013; HanDeDict, 2011; Idiomdatenbank, 2007].

German
- wieder da, wo wir angefangen haben (still there, where we started/began)
- noch einmal von vorne beginnen (to start again from scratch)
- wieder bei Null anfangen (to start again from zero)
- zurück zum Start (back to the start)
- auf Null zurück (back to zero)
- wieder zum Anfang zurückgehen (to go back to the beginning/start)
- alles auf Null (all to zero)
- zurückgehen (to go back)
- zurück auf Platz eins sein (to be back at/place one)

French
- revenir à la case “départ” (come back to the box “start”)
- ramener à la case “départ” (come back to the box “start”)
- en crée de nouveau (to create something anew)
- retourner à la case départ (come back to the box “start”)
- être au point mort (to be at a deadlock)
- remettre le compteur à zéro (set the counter back to zero)
- repandre la case départ (to continue from the box “start”)
- revenir au point du départ (come back to the back)
- repartir à zéro (to start again from zero)
- faire table rase (start afresh/start anew/start over/turn a new leaf/start all over again)

Spanish
- estar en el punto cero (to be at point zero)
- partir de cero (start/begin from zero)
- empezar desde cero (start/begin from zero)
- una vuelta al principio (a turn to the beginning)
- hasta el nivel cero (by level zero)
- comenzar de nuevo (start again anew)
- devolver al punto de partida (return to the point of origin/to the starting point)

Portuguese
- de volta ao ponto de partida (back to the point of origin/to the starting point)
- estar no ponto zero (to be at point zero)
- regressar ao ponto de partida (to go back to point zero)
- estar de volta à estaca zero (be at point zero)
- voltar à estaca zero (turn back to point zero)
- recomeçar do princípio (start again from scratch)
- criação de nova (create anew)

- voltar ao principio (return back to the origin/the start)
- voltar a como era antes (return to how it was)
- até au nível zero (to level zero)

Italian
- ricominciare da capo/daccapo (start/begin anew)
- cominciare da zero (start from zero)
- ricominciare da zero (start from zero)
- partire dall’inizio (start from the beginning)
- tornare al punto di partenza (start from the starting point)

Chinese
- 另起炉灶 (lingquliuzao) (to set up a separate kitchen, start from scratch) (`炉灶` kitchen range, kitchen stove)

Swedish
- tillbaka till grunden (going back to the ground)
- börja om från början (start from the beginning)
- tillbaka till utgångspunkten (go back to the point of exit)
- tillbaka vid utgångspunkten (go back to the point of exit)
- börjar från noll (start from zero)

The first observation from the data teaches that it is possible to paraphrase going back to square one via other abstract forms or idioms. To go back to square one has the same meaning of and can thus be translated into, inter alia, to start from scratch, to make a fresh start, to go back to point zero, to set up a separate kitchen. Interestingly, some of these expressions could only be found in the process of translating the metaphors from other languages into English. Within the limits of the tools and reference consulted, I have found no repository that comprehensively reports all possible idioms for the same concept in one language or more than one.

In six of the seven languages presented, there are regular semantic patterns (at least regular in comparison to the semantic elements of the original English form considered). These patterns are: The expressions contain a shape and a number. Going back to square one is mostly translated into start from zero or be at point zero. The choice of picking up Romance, Neogermanic and Oriental languages at the same time makes the comparison less vulnerable to remarks on linguistic inheritance.

Only in Chinese there is, in comparison to the other forms in the other languages, a completely different word, ‘kitchen stove’ (which relates to the English idiomatic expression to set up a separate kitchen).

So, to answer the questions posed before the cross-linguistic analysis, it seems that the presence of a shape in the source domain is not binding for defining the concept of start/begin in the target. Relating this finding to the above presented definition of metaphor of shaping, the shaping or process of the form coming into being is thus located in the source.

We may not find a form in all idioms in all languages, but we find in almost all of them a numbering, which enables us to state that “Starting something involves going back or start either from zero or from one.”

The comparison shows, by linguistic comparison alone, a recurrent use of shapes and numbers for the same expression among different languages. But the translation with shapes is
not the only translation possible. This gives indication that the concept of shape, let it be square or point, might not be so indicative for the meaning of the idiom. If so, we are confronted with another metaphor of shaping and not of shape.

### 4.2 Sense and Meaning

In the language of current research on formal conceptual mapping, lexical entries sense the same lexical sense. I am referring here to the OntoLex terminology as adopted by John McCrae in his diagrammatic rendering of semiotics.ow ling (published on ontolex w3.org).

According to the scheme, a LexicalEntry denotes an OntologyEntity, expresses a Concept or Synset and senses a LexicalSense. The LexicalSense refers to the OntologyEntity and the Concept or Synset conceptualizes it. From this Concept, we can derive the semiotic Meaning.

The scheme is still object of investigation by the W3C members, so that a final word on the definiteness of the model might be inappropriate but group members seem to agree that a distinction is not sense and meaning in ontological and lexico-semantic formalizations represents added value. (last update of the model checked on: April 19th, 2013).

The interlingual analysis of the idiom going back to square one shows that: (a) the lexeme\(^ {10} \) can be paraphrased in many other lexemes in the same language; (b) the lexeme exists in different languages with different semantic forms including space number zero or number one; and (c) it always conveys the same meaning.

Following the suggested OntoLex model, I want at this point claim that the target domain of an idiom entails the idiom’s concept, but for the idiom to have a meaning, it needs a sense, contained in the source domain.

The idiom to go back to square one denotes the concept or OntologyEntity beneath ‘start/Begin’. This is an encyclopedic truth, something we can learn from thesauri and dictionaries. Cross-linguistic analysis alone tells us that a process of starting or beginning something in conjunction with this idiom can only happen when it entails the sense of going back to something, which is also visualized with a space number zero or number one.

‘Sense’ is here used not with the connotation of ‘meaning’ (as in the case of a polysysemous word), but for indicating an adjunct which reinforces and fulfills the lexical meaning. The interpretation of the term does not dissociate from its standard definition.

Lexical senses, like lexical entries, are language-specific. In the comparison, I have found different lexical entries for one language among different languages. Nevertheless, these languages also share the same lexical sense for the same lexical entry.

The Chinese idiom for going back to square one, 另起炉灶 (lìngqūlúzào), has been the only case with the most divergent semantic formation (its English correspondent: to set up a separate kitchen).

If we argue once again (contextual use can further prove the validity of this claim) that the metaphor is not of shape, but of shaping, than ‘kitchen’ can stand for more than a single room, but also for factory, business, facility, activity, productivity, event. The same can be claimed for all the other metaphors in the list. ‘Head’ can then stand for ‘idea’ or ‘behavior’ or ‘trend’. To reduce the fuzziness of these implied conceptualization, [Hobbs, 1992]’s model might become handy as well (the same model that declares why John is an elephant for his clumsiness and not for his weight).

The conceptual and contextual and crosslinguistic analysis together trigger discussion not just in the domain of conceptual encoding, but also in creativity and in the way language users indulge or procrastinate new ideas.

### 5 Conclusions and Future Work

Metaphors are at the backbone of everyday communication. We utter almost six metaphors a minute, once said Geary James in a TED talk\(^ {11} \), and yet it is unclear where they sit in the brain.

Different theories have been suggested, the research is ongoing.

From the linguistic point of view, metaphors are interesting constructs that have been synthetized in a source and a target domain. For some scholars, this is a rather imprecise distinction which leaves little attention to cultural and contextual differences; for others, it is a feasible way to extract concepts automatically.

I pick up two case studies, vicious/virtuous circle and going back to square one, and analyze their crosslinguistic contextual behavior (with an emphasis of the latter for the form with ‘square’ and an emphasis on the monolingual corpus-based analysis for ‘circle’).

In this research, I work with metaphors and shape forms in the source domain. In the best of my literature knowledge I have so far encountered no cases of the two studied together. I make a distinction between metaphors of shape and metaphors of shaping, because I suppose that not all shapes are equally important in the understanding of a metaphor that entails them, if they are relevant at all. In metaphors of shape, the precise visualization of the form is necessary to understand the meaning of the expression. In metaphors of shaping, the geometric forms, its visualization and representation are relevant to the final meaning of the form. The shape is structural to the linguistic form, not to its meaning, but it gives indication of the spatial/temporal settings in which the event occurs.

The qualitative analysis so far enables to tell that both forms are metaphors of shaping and not of shape, that in none of them the form is relevant. But the research beside this study has also showed that the are complex forms that might be of shape or of shaping according to the context, so that a further investigation in this direction is needed.

Another remark is on the literal meaning of the metaphors, which enables to detect deeper “suppliers of meaning” or senses, that otherwise would go lost in the commonsense use of the form.

The way metaphors behave, the concepts they trigger, the place in the brain they claim, the process of creativity they

\(^ {10} \) The idiom is defined here as lexeme or lexical entity; OntoLex reference

\(^ {11} \) blog.ted.com/2009/12/18/metaphorically/
References


[Ramachandran & Hubbard, 2011] Vilayanur R. Ramachandran & Edward M. Hubbard. The Emergence of Qualia,


Semantic Clues for Novel Metaphor Generator

Rafal Rzepka†, Pawel Dybala‡, Koichi Sayama‡ and Kenji Araki†
†Hokkaido University, Sapporo, Japan
{karura,araki}@media.eng.hokudai.ac.jp
‡Otaru University of Commerce, Otaru, Japan
{paweldybala,sayama}@res.otaru-uc.ac.jp

Abstract

In this position paper we introduce our early findings drawn from a new study on an over 84,000 entries collection of metaphorical expressions manually extracted from Japanese literature. Such thoroughly chosen data assure higher standards when figurative speech is analysed from the generation perspective and allow to develop an algorithm for explaining difficult or new notions. Such function is going to be implemented in our artificial tutor project and here we describe a method for generating similes. We also show some statistical discoveries in the semantic level that could be useful knowledge for a dialog system with extended explanation capabilities.

1 Introduction

Figurative speech is one of the most spectacular thought processes used in human communication. If we need to explain some difficult word, to delicately suggest or emphasize something, we often use metaphors. Good teachers and famous poets have been using imaginative examples to help us understand things, abstract phenomena and complex or straightforward emotions. Although human beings usually have no problems with creating such examples, choosing an understandable metaphor that will trigger somebody’s imagination is a difficult cognitive process [Carbonell, 1982], [Mcglone, 1996]. The most famous theories on how we understand metaphors are the categorization view [Glucksberg, 2001], the comparison view [Gentner, 1983] and three hybrid views – the conventionality view [Bowdle and Gentner, 2004], the aptness view [Jones and Estes, 2005] and the interpretive diversity view [Utsumi and Kuwabara, 2005]. This paper’s goal is not to describe computational models for these methods with their pros and cons but to support researchers working on figurative speech generation as an addition to the existing theories and prepare ourselves to implement them as a part of explanation module for our artificial tutor project [Mazur et al., 2012]. We used more than 12,000 of 30,000 metaphorical expressions (similes) gathered by Onai [Onai, 2005] which allows for a new approach for computing or analysing allegorical utterances in natural language interfaces. First, we show statistical data about words usage, then we propose a simple similes generating algorithm and finally describe preliminary experiments for setting understandability threshold. We conclude with discussion about possible usage of the new data set and introduce our plans for the fuller usage of the data.

2 Data Analysis

2.1 Onai’s Dictionary

Metaphors used in this study were acquired from Onai’s Great Dictionary of Japanese Metaphorical and Synonymical Expressions [Onai, 2005]. The dictionary contains metaphors selected from Japanese modern literature and Japanese translations of foreign works. The dictionary contains approximately 30,000 metaphorical entries, each of which includes:

- headline, i.e. a word or phrase used to look up metaphors.
- sometimes – sub-headlines, i.e. words or phrases similar to the headline
- index – a phrase that was actually used in that particular metaphor (or its semantic equivalent)
- metaphor – actual metaphor example
- source – reference to the literature work from which the metaphor was selected.

According to the author, the dictionary was compiled to assist in finding interesting and somewhat sophisticated expressions that can be used instead of common phrases. If, as in the example entry in Table 2.1, one needs to find an unusual expression for “a rough woman”, first he would have to query the word “woman” (headline), then search for the particular expression in the index and finally check the actual metaphor example.

2.2 Semantic Characteristics

To our best knowledge, the data we used is the biggest digitized collection of Japanese metaphorical expressions and can be analysed from various angles. For the first trials with generation we have chosen the simplest and the most popular metaphorical figure of speech – a simile. A simile differs from a metaphor in that the latter compares two unlike things by saying that the one thing is the other thing, while simile directly compares two things through some connective, usually
“like”, “as” or by specific verbs like “resembles”. In order to select similes from our data set, we used a manually created set of such words (marks) used in Japanese. This allowed us to retrieve 12,214 similes, on which we performed some statistical tests. By using JUMAN morphological parser\(^1\) we have separated and ranked 3,752 unique part of speech patterns that can be helpful while generating figurative expressions. Dependency parser KNP’s dictionary\(^2\) and semantic role tagger ASA\(^3\) were then used in order to rank most popular categories and words needed to set weights for deciding on the best simile candidate in the generation process. Most frequent semantic categories characteristic to figurative speech were colors, shapes, and patterns. The words with highest frequency are shown in Table 2 (grouped by part of speech). For comparison, semantic categories characteristic to random web text (3,500 sentences from a blog corpus) were mostly places, currencies, names, dates, organizations, education and the words most characteristic to a random web text were as follows.

Nouns:
- learning, media, bad, work, information, method, enterprise, understanding, company, strength, area, necessity, relationship, usage, utilization, direction, United States, system, administration, thought, two, city, money, district, caution

Verbs:
- to be visible, to know, to be divided

Adjectives:
- individual, many, old

Further semantic analysis data can broaden the system’s knowledge and become also helpful for recognizing figurative speech because when understanding users’ utterances as metaphorical or idiomatic expression they need to be processed by using different comprehension strategies.

### Table 1: An example entry in Onai’s dictionary

<table>
<thead>
<tr>
<th>Japanese Transcription</th>
<th>Index</th>
<th>Metaphor</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>onna; josei</td>
<td>hageshii onna</td>
<td>hi no you ni hageshii onna</td>
<td>Jakuchou Setouchi</td>
</tr>
<tr>
<td>a woman; a lady</td>
<td>rough woman</td>
<td>woman rough like a fire</td>
<td>(a book by Jakuchou Setouchi)</td>
</tr>
</tbody>
</table>

### Table 2: Top 10 nouns, verbs and adjectives out of 10,658 morphological tokens found in corpus (similes only).

<table>
<thead>
<tr>
<th>Nouns</th>
<th>Verbs</th>
<th>Adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>eye (524)</td>
<td>to look like (236)</td>
<td>white (265)</td>
</tr>
<tr>
<td>voice (437)</td>
<td>to flow (204)</td>
<td>black (156)</td>
</tr>
<tr>
<td>water (338)</td>
<td>to shine (197)</td>
<td>beautiful (138)</td>
</tr>
<tr>
<td>sound (330)</td>
<td>to stand (182)</td>
<td>cold (135)</td>
</tr>
<tr>
<td>face (325)</td>
<td>to look (169)</td>
<td>heavy (107)</td>
</tr>
<tr>
<td>heart (309)</td>
<td>to fall down (161)</td>
<td>dark (101)</td>
</tr>
<tr>
<td>breast (282)</td>
<td>to put out (146)</td>
<td>sharp (101)</td>
</tr>
<tr>
<td>light (265)</td>
<td>to go (145)</td>
<td>big (96)</td>
</tr>
<tr>
<td>sky (235)</td>
<td>to move (139)</td>
<td>small (93)</td>
</tr>
<tr>
<td>head (218)</td>
<td>to feel (137)</td>
<td>detailed (86)</td>
</tr>
</tbody>
</table>

In order to examine the applicability of a generation task, the experimenter must conduct a metaphor generation task with a huge number of concepts, therefore Abe et al. used Japanese newspaper corpus as a base for their language model. Researchers also use the Web as a source for their models [Veale and Hao, 2007][Masui et al., 2010] and utilize the latest thesauri and ontologies as WordNet [Miller, 1995] to build sophisticated generation algorithms [Huang et al., 2013]. Numerous examples extracted from Onai’s dictionary could be helpful for all existing approaches. Therefore we are planning to test most popular approaches in nearest future. For the basic preparations we have used the 12,214 similes mentioned in the previous section. Currently we are working on Ortony’s salience imbalance theory [Ortony, 1979] which predicts possible source-target shared attributes and their positions in each ranking. Together, these concepts imply that low-high topic-source pairings should cause increases in salience of topic attributes. On Figure 1 we show the idea of two lists of attributes that describe a word in an order of occurrences. So “sweet honey” is more natural than “sweet voice” but the same adjective can describe both nouns. However, as Ortony’s theory suggests, if two adjectives are from the top or bottom of the lists (distance between them increases), it is less likely that they can form an apt simile. We propose a method for calculating this distance in the following section.

#### 3.1 Toward Calculating The Salience Imbalance

To observe which attributes could be used for a ground combining a source and a target (as “strong” in man strong as a bear) we experimented with two attribute lists. Preliminary tests with different data sets suggested that it is fairly probable to find common attributes between distant source and target nouns using the Japanese case frames database [Kawahara and Kurohashi, 2001], but as it is mostly verb centered, we also created attributes lists for nouns used in our

---


\(^2\)Japanese Dependency and Case Structure Analyzer KNP 4.0: http://nlp.iist.i.kyoto-u.ac.jp/index.php?KNP

\(^3\)ASA 1.0 - Semantic Role Tagger for Japanese Language: http://cl.it.okayama-u.ac.jp/study/project/asa/
trials (an example is shown in Figure 1). We concentrate on web-based textual resources because we aim at agent’s dialog capabilities for using figurative speech mainly among Japanese teenagers studying English. Newspapers have not many metaphorical expressions and freely available corpus for Japanese literature consists mostly of old books with old written language. We chose one simple pattern: Source - Mark - Ground - Target, which in English would be Target - Ground - Mark - Source as in Train small like a matchbox. The algorithm we created uses original source and target pair and compares lists of semantically related phrases retrieved from the web. For instance associations list for fire contains phrases as “to set”, “to put out”, “weak”, “fast” or “extinguisher”. When we input 100 random phrases according to a chosen pattern, only 16 source-target pairs had common attributes and this is because there are less adjectives than verbs in the data. We have calculated the positions on the attribute lists that are sorted according to the frequency in the web corpus so to set fire is on the top to carry fire is closer to the bottom. We have calculated the distance value between common attributes using the following formula.

\[
distance = \frac{\text{SourcePosition}}{\text{TotalSourceAttr}} \cdot 100 - \frac{\text{TargetPosition}}{\text{TotalTargetAttr}} \cdot 100
\]

For example, from the metaphor “train small as a matchbox”, the system first extracts “matchbox” (source) “train” (target) and “small” (ground). Next, the rankings of attributes of “train” and “matchbox” are extracted from the case frames database, and the average position of “small” in each ranking is checked. Finally, the system (after multiplying each position by 100 to avoid dealing with very small fractions) calculates the difference of grounds position in these two rankings.

- Metaphor: Train small as a matchbox

- Source: matchbox
- Target: train
- Ground: small
- Total source attributes: 64
- Total target attributes: 7444
- Ground position in source attributes ranking (SourcePosition): 21
- Ground position in target attributes ranking (TargetPosition): 5088

\[
distance = \frac{21}{64} \cdot 100 - \frac{5088}{7444} \cdot 100 = 35
\]

The results are shown in Tables 3 and 4.

### 3.2 Preliminary Evaluation

As we plan to use the generator for a English tutoring dialog system, we need to be sure that the example chosen by computer is not too metaphorical and difficult to understand. To set the thresholds we performed a preliminary evaluation experiment and asked 5 Japanese language speakers (males only) to evaluate phrases used in the previous section. The subjects were asked to rate understandability and aptness on the 5 points scale. The results (see Table 3) show that human evaluation averages for both aspects have set a borderline for distinguishing good similes (“voice sweet as honey” or “hand cold as ice”) from lesser ones as “hand delicate as woman” which is a repetition abbreviation of hand delicate as woman’s hand.

By using human made examples from Onai’s dictionary we were able to set an average distance \((D)\) to 20. Because setting only distance was generating too many candidates, we...
Table 3: Sixteen metaphors which had common Ground in a Kyoto Case Frames-based attributes lists. Sorted in order in distance which is the difference between attribute usualness values calculated from Ground position in both lists.

<table>
<thead>
<tr>
<th>simile</th>
<th>average understandability</th>
<th>average aptness</th>
<th>difference between understandability and aptness</th>
<th>difference between attribute positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>train small as a matchbox</td>
<td>4.8</td>
<td>4.2</td>
<td>0.6</td>
<td>35</td>
</tr>
<tr>
<td>skin thin as paper</td>
<td>3.966</td>
<td>2.8</td>
<td>1.166</td>
<td>0.9</td>
</tr>
<tr>
<td>hand cold as ice</td>
<td>4.4</td>
<td>4.4</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>woman rough as fire</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>hand delicate as woman</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>skin white as bleached</td>
<td>5</td>
<td>4.6</td>
<td>0.4</td>
<td>5</td>
</tr>
<tr>
<td>hand cold as a fish</td>
<td>2.6</td>
<td>2.6</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>eyes small as elephant</td>
<td>2.2</td>
<td>2.8</td>
<td>-0.6</td>
<td>4</td>
</tr>
<tr>
<td>voice sweet as honey</td>
<td>4.2</td>
<td>3.8</td>
<td>0.4</td>
<td>33.3</td>
</tr>
<tr>
<td>eye sharp like being on the watch</td>
<td>3.6</td>
<td>3</td>
<td>0.6</td>
<td>23</td>
</tr>
<tr>
<td>lips thin as a blade</td>
<td>3.4</td>
<td>3.48</td>
<td>-0.08</td>
<td>31</td>
</tr>
<tr>
<td>wife young as a daughter</td>
<td>4.4</td>
<td>3.4</td>
<td>1</td>
<td>76</td>
</tr>
<tr>
<td>lips red as blood</td>
<td>4.8</td>
<td>4.4</td>
<td>0.4</td>
<td>1</td>
</tr>
<tr>
<td>oysters big as zori thongs</td>
<td>4</td>
<td>3.4</td>
<td>0.6</td>
<td>23</td>
</tr>
<tr>
<td>sleep deep as death</td>
<td>5</td>
<td>4.8</td>
<td>0.2</td>
<td>1</td>
</tr>
<tr>
<td>face long as a horse</td>
<td>4</td>
<td>3.4</td>
<td>0.6</td>
<td>49.5</td>
</tr>
<tr>
<td><strong>averages for thresholds:</strong></td>
<td><strong>3.96</strong></td>
<td><strong>3.63</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Results for understandability and aptness evaluation experiment. Lines in gray show similes which had both understandability and aptness higher than the averages.
have performed multiple additional experiments with different parameters to see which conditions are helping and which are not. Results of one strict set of conditions is shown in Table 5. In most cases, if semi-correct results are counted as positive, the newly generated similes were significantly better that a random generation, but further filtering semantically strange outputs is needed.

4 Conclusions and Future Work

In this short paper we have introduced new possibilities for figurative speech generation by using a new vast collection of Japanese metaphorical expressions. Because this is an early stage of our study, we have performed only preliminary tests and experiments in order to get a grasp of which tools and other repositories can be combined before we start implementing the data to known theories about human's ability to use examples while explaining physical and abstract objects. We have already started to work with more simile patterns, also including verbs to fully utilize Kyoto Frames database. We are experimenting with N-gram frequencies of target, ground and source triplets to create vectors which should help us discover more statistical dependencies. We are also testing WordNet and ConceptNet [Liu and Singh, 2004] as a source for further calculation of semantic dependencies and show the latest progress during the workshop.

References


<table>
<thead>
<tr>
<th>English translation</th>
<th>Japanese in Roman letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lips as red as they were licking blood</td>
<td>Chi-o nameru you-na akai kuchibiru</td>
</tr>
<tr>
<td>Lip as thin as somebody press a blade against it</td>
<td>Ha-o oshitsukeru you-na usui kuchibiru</td>
</tr>
<tr>
<td>Voice so sweet as somebody was sucking up honey</td>
<td>Mitsu-wo suiageru you-na amai koe</td>
</tr>
<tr>
<td>Wife as young as somebody was talking to daughter</td>
<td>Musume-ni hanasu you-na wakai tsuma</td>
</tr>
<tr>
<td>Wife so young that resembles daughter</td>
<td>Musume-ni niru you-na wakai tsuma</td>
</tr>
<tr>
<td>Woman so rough like she was dropping fire</td>
<td>Hi-o otosu you-na hageshii onna</td>
</tr>
<tr>
<td>Sleep as deep as somebody was avoiding death</td>
<td>Shi-o kaihi-suru you-na fukai nemuri</td>
</tr>
<tr>
<td>Hand as cold as somebody was biting ice</td>
<td>Koori-o kamu youna tsumetai te</td>
</tr>
<tr>
<td>Hand as cold as somebody was putting it into ice</td>
<td>Koori-ni ireru you-na tsumetai te</td>
</tr>
<tr>
<td>Skin as thin as somebody was cutting off paper</td>
<td>Kami-o kiritoru you-na usui hifu</td>
</tr>
<tr>
<td>Skin as thin as somebody was peeling off paper</td>
<td>Kami-o hagu you-na usui hifu</td>
</tr>
<tr>
<td>Skin as thin as somebody was scratching off paper</td>
<td>Kami-o hakkaku you-na usui hifu</td>
</tr>
<tr>
<td>Skin as thin as somebody could stick paper (to it)</td>
<td>Kami-ni haritsukeru you-na usui hifu</td>
</tr>
<tr>
<td>Face as long as somebody was being put on a horse (back)</td>
<td>Uma-ni noseru you-na nagai kao</td>
</tr>
<tr>
<td>Face as long as somebody was aiming at a horse</td>
<td>Uma-ni ateru you-na nagai kao</td>
</tr>
<tr>
<td>Face as long as somebody was separated from a horse</td>
<td>Uma-kara hanasu you-na nagai kao</td>
</tr>
<tr>
<td>Wife as young as somebody was passing to daughter</td>
<td>Musume-ni watasu you-na wakai tsuma</td>
</tr>
<tr>
<td>Wife so young that you could get used to daughter</td>
<td>Musume-ni nareru you-na wakai tsuma</td>
</tr>
<tr>
<td>Wife so young that she could take away daughter</td>
<td>Musume-wo ubau you-na wakai tsuma</td>
</tr>
<tr>
<td>Wife so young that you could find in daughter</td>
<td>Musume-ni mitsukeru you-na wakai tsuma</td>
</tr>
<tr>
<td>Wife so young that you could be worried about daughter</td>
<td>Musume-ni kizukau you-na wakai tsuma</td>
</tr>
<tr>
<td>Wife so young that you could let her go to daughter</td>
<td>Musume-ni hanasu you-na wakai tsuma</td>
</tr>
</tbody>
</table>

Table 5: Examples for an Kyoto frames-based generation experiment for candidates with the same particles and grounds, where there were more than 30 verb-source and 30 grounds in Ameba corpus. Conditions for salience imbalance distance values were $D < 100$ and $D > 5$. First group of examples were classified by authors as correct, second as semi-correct and third as incorrect, although there was no agreement in case of few examples from the last group suggesting that imagination plays a big role in evaluating novel metaphors.