Towards comparable cognitive creative systems. Two case studies and a general approach based on cognitive processing, knowledge acquisition and evaluation with human creativity tests.

Ana-Maria Olteţeanu
Cognitive Systems, Universität Bremen
Bremen, Germany
amoodu@informatik.uni-bremen.de

Abstract

Creative cognitive systems are rarely assessed with the same tools as human creativity. In this paper, a general approach is proposed for building more such comparable systems. The importance of using cognitively viable processes, cognitive knowledge and evaluation when implementing creative problem-solving systems is emphasized. Two case studies of artificial cognitive systems evaluated with human creativity tests are reviewed. A general approach is put forward. The applicability of this general approach to other creativity tests and artificial cognitive systems, together with ways of performing cognitive knowledge acquisition for these systems are then explored.

1 Introduction

Imagine you are in your house, needing to solve a particular problem which involves specific objects. You need a particular tool, object (e.g. a piece of string) or recipe ingredient (e.g. mince meat), but you don't have it in the house. You would like to solve your problem using a different tool, object or ingredient, and have a system show you what you could use instead (e.g. dental floss to replace the piece of string, aubergine to replace the mince meat - depending on task and recipe context).

Or envision a situation in which you ran out of ideas on how to solve a particular problem. You feel stuck and would like to think of a different approach. You would like the help of a system that could inspire you, operating in a cognitive manner, and showing you items (websites, research articles, excerpts from encyclopedias, music, films, photos, formulas) which could trigger for you new associations, new ways of solving that problem.

Such assistive systems would need to be capable of creative problem-solving and of presenting their results in a manner that can easily be used as input by humans. They would need to be endowed with cognitive knowledge acquisition, and types of knowledge processing which are akin to those used by humans in their creative problem-solving.

Computational creativity is a strongly emerging field in Artificial Intelligence, with application systems ranging from poetry [Colton *et al.*, 2012] and painting [Colton, 2012] to mathematics [Lenat, 1976].Empirical tests of human creativity and creative problem-solving do exist. However, computational creativity systems can rarely be assessed in a comparable manner, i.e. by using human creativity tests. This has as consequence the fact that not many artificial cognitive systems which can be used as cognitive models exist (e.g. [Hélie and Sun, 2010]), and thus not as much progress is made as it could be in terms of understanding the cognitive bases of the creative process, and in building artificial cognitive creative systems.

This paper argues that (i) artificial cognitive systems can be used to shed light on the human creative process and (ii) knowledge obtained from creativity tests can be used to inform and evaluate artificial cognitive systems, if more work is done on artificial cognitive systems which can be assessed with human creativity tests. To support this claim, this paper presents two case studies of systems which can solve human creativity tests, briefly describing how cognitive knowledge was acquired and organized for such systems, the processes used and how the systems where evaluated compared to their human counterparts. A general approach for building more such systems for other creative tasks is presented, together with the types of knowledge, knowledge acquisition, processes and types of evaluation which can be used. The benefits of bridging this gap are then shown in terms of (a) new relations which can be observed or studied from a cognitive modeling paradigm and (b) knowledge and data obtained from human creativity tests which can then be used to inform artificial cognitive systems.

The rest of this paper is structured as follows. Section 2 describes the differences between computational creativity evaluation methods and human creative evaluation. Cognitive processes relevant when implementing artificial creative cognitive systems are discussed in Section 3. Sections 4.1 and 4.2 briefly describe two case studies of artificial cognitive systems which can give comparable results to humans in creativity tests, together with the knowledge acquisition, cognitive processing and evaluation of these systems. Section 5 presents a general approach when aiming to make artificial cognitive systems yield comparable results to humans in creativity tests. The applicability of this approach to other creativity tests is discussed in Section 6. A short description of how cognitive knowledge acquisition can be performed using creativity tests is provided in section 7. A short discussion and conclusion is presented in Section 8.

2 Computational creativity evaluation versus human creativity tests

Computational creativity is evaluated in various ways and it has been debated what the evaluation process for computational creativity systems should be [Wiggins, 2001; Ritchie, 2001]. Some authors aim to produce artefacts through their systems, and would like them to be assessed as having comparable creativity as that of humans. Others define creativity in terms of process [Boden, 2003], not aiming for an implementation. Ritchie [2001] proposed an assessment of computational creative systems which takes into account the inspiring set – a union of implicit and explicit knowledge – formalizing fourtheen criteria of evaluation, centered around typicality, quality and novelty. Pease, Winterstein and Colton [2001] proposed an evaluation which takes into account the input, output and process of the creative system, based on measures of novelty, quality and process. The FACE model [Colton *et al.*, 2011] describes creative acts as tuples of generative acts, including items and methods of producing items in four categories: concepts, concept expressions, aesthetic measures and framing information. The IDEA model [Colton *et al.*, 2011] uses the notion of impact a creation has, rather than that of creation value, and suggests six stages of development of a computational creativity system, based on the difference between the information given as knowledge to the system and the artefacts it generates. Other authors use human evaluation (e.g. Williams and McOwan [2014]), asking users to rate the products of computational creativity systems, or to choose words which appropriately describe their reaction to those products.

Though the field of computational creativity evaluation has made significant progress, none of these types of evaluation shines a bright light on cognitive creative processing.

On the other hand, human creative problem-solving or processes considered to take part in it (like divergent thought) are evaluated with various tasks. Some of these are the following:

- The Remote Associates Test [Mednick and Mednick, 1971];
- The Alternative Uses Test [Guilford, 1967];
- The Torrance Creativity Tests (reviewed by Kim [2006]);
- The Wallach-Kogan test [Wallach and Kogan, 1965];
- Insight tests [Maier, 1931; Duncker, 1945]

The human responses in such tasks are generally evaluated in terms of various groups of the following metrics. Some of these metrics assess a particular expected correct answer, others are used for assessing open-ended answers.

1. achieving or not achieving a solution (this metric is particularly useful for hard insight tasks);
2. response time for a particular solution;
3. difficulty of solving a particular problem item as a percentage of the population solving it;
4. fluency - in open-ended tests, measures how many different items the participant has come up with as an answer);
5. flexibility - measures how many semantically different domains the answers to a particular item cover;
6. elaboration - assesses the amount of detail contained in the various answers;
7. originality - responses given by a small percentage of the participants are rated as unusual;
8. novelty - human judges are asked to assess answers given by human participants on a novelty scale (used for the Alternative Uses Test [Guilford, 1967], the Wallach-Kogan test [Wallach and Kogan, 1965]). One can then check for the validity of the novelty judgement by exploring the agreement between the assessments of the various judges.

Furthermore, the literature addressing these tests sheds insight into the cognitive processing behind such tasks.

3 The importance of cognitive processing in creative problem-solving systems

Various cognitive processes are said to account for creative problem-solving, with one of the most modeled processes being analogy (with models like ANALOGY [Evans, 1964], MAC/FAC [Forbus *et al.*, 1995], LISA [Hummel and Holyoak, 1997], STAR [Halford *et al.*, 1994], Copycat [Hofstadter *et al.*, 1994], etc.). Another creative process which has recently gathered investigative interest is that of conceptual blending [Fauconnier and Turner, 1998]. Various theories account for insight problem-solving [Batchelder and Alexander, 2012]. In order to build artificial cognitive creative agents, one can use this literature to build computational mechanisms akin to the cognitive processes involved in solving creative tasks. In the author's opinion [Olteţeanu, 2014], the cognitive processes of association, use of similarity, structure, and re-representation are to be always kept in mind.

Associations are important in creative problem-solving due to their ability to bring new material into the problem space for the solver. Creative problem spaces are ambiguous and therefore can benefit from fluidity. Associations can easily be made based on **similarity** or **context**. Associations can be made by context, as encountering certain items (a, b, c, d) constantly together, generally produces the cognitive result of triggering the other items (c, d) when some of the items (a, b) have been shown. Associations can be made by similarity, as similarity of features might imply similarity of affordance (e.g. if you know you can kick footballs, you might want to try playing in the same way with any spherical object of similar weight and material).

All such processes can help bring more useful items into the problem space and allow **re-representation**. Knowledge of **structure** can allow structure-mapping into different domains [Gentner, 1983], replacement of structure parts [Olteţeanu and Falomir, 2015b], navigation between similar structures, and structure-based operations (merging, overlap, removal of unnecessary parts, etc.)[Olteţeanu, 2014]. This allows further ways of using old knowledge creatively in order to produce new, useful and interesting knowledge.

Such cognitive processes need to be replicated in artificial cognitive systems in order to: a) enable them to assist the humans in processes of re-representation and creative problem-solving, and b) explain their own creative productions in ways which make sense to humans. An initial requirement in replicating and refining such processes is the ability to build systems capable to give results comparable to humans in various creativity tests.

4 Case studies

This section summarizes two case studies of cognitive systems which yield results comparable to human participants in human creativity tests. Section 4.1 presents the human Remote Associates Test and a cognitive system which can give comparable answers from the literature [Olteţeanu and Falomir, 2015a]. Section 4.2 presents the Alternative Uses Test and a system capable of giving similar answers to humans to this test [Olteţeanu and Falomir, 2015b].

4.1 Case 1 - A Remote Associates Test solver

This section reviews: (i) the Remote Associates Test, (ii) the cognitive knowledge acquisition and process of the computational solver and (iii) the evaluation with human data.

The Remote Associates Test

The Remote Associates Test (RAT) by Mednick and Mednick [1971] is a creativity test in which participants are given three word items and required to produce a fourth, which is associates with all three of them in some way. For example, the words CREAM, SKATE and WATER are given. A correct answer to this query would be the word ICE. Various types of RAT can be distinguished, depending on the type of associations at play [Worthen and Clark, 1971].

The RAT has been widely used as an empirical measurement of creativity [Ansburg, 2000; Dorfman *et al.*, 1996] and translated into multiple languages [Baba, 1982; Hamilton, 1982; Nevo and Levin, 1978; Chermahini *et al.*, 2012].

The type of data acquired after empirical testing includes the percentage of participants solving a particular test item and response speed. Response times for different amounts of allocated solving time have been obtained in the literature (for 7s, 15s, 30s), with useful sources of normative data being available [Bowden and Jung-Beeman, 2003].

Description of the knowledge acquisition and process for the artificial cognitive system solver

A computational RAT solver (comRAT) was presented in the literature by Olteţeanu and Falomir [2014; 2015a], using language data and a convergence process, which will be further briefly summarized. The language data used consisted of the $1,048,720$ most frequent 2-grams from the Corpus of Contemporary American English[1]. Using the UCREL CLAWS7 Tagset[2], only $205,602$ 2-grams with relevant tags for the task were kept. The comRAT system learned all the unique $Concepts$ of the 2-grams, and organized bidirectional $Links$ between $Concepts$ which appeared together in an $Expression$.

When 3-word RAT queries were given to the system, if known, the 3 $Concepts$ were activated, together with the $Links$ and $Concepts$ attatched to them. The $Concepts$ activated from most sides were considered as potential answers. The activation of the initial Concepts thus converged upon possible answers for RAT queries.

Description of evaluation with human data

The comRAT system was given the 144 items from the Bowden and Jung-Beeman [2003] normative data to solve, thus being tested with the exact same queries as humans. Without using any frequency data, the computational RAT found the correct answer provided in the normative data for 47 out of the 48 items for which it had known all 3 initial Expression items. The system solved another 17 queries for which it only knew 2 initial items. For another more than 20 cases, the system came up with other plausible answers, which were not identical to the correct answers provided by the normative data.

Using frequency data, the probability of the system to find an answer has been found to correlate with the difficulty of the RAT queries for humans. Difficulty of query for humans was understood to be represented in the normative data by human response time and percentage of participants solving the query. A significant moderate correlation between difficulty of query for humans and comRAT's probability of finding an answer was observed.

The authors mention that the comRAT solver can be used to generate new RAT queries. RAT for human participants can thus be generated controlling for semantic tag, position in the 2-gram of the given and answer word, frequency. Some of these queries can have multiple possible answers. In this case, data about probability of answer from the computational system can be used to check whether (i) answer preference in humans is due consistently to a higher frequency of the query item-answer pair, or (ii) appreciation for a particular answer is higher when the frequency is lower.

4.2 Case 2 - An Alternative Uses Test solver

This section reviews: (i) the Alternative Uses Test, (ii) the knowledge organization and process of the computational solver and (iii) the evaluation of the system, comparable to evaluation of human answers to the Alternative Uses Test.

The Alternative Uses Test

The Alternative Uses Test [Guilford, 1967] takes the following form: participants are given the name of an object item (e.g. $Brick$), and asked to come up with as many different uses as they can for that item, in a set amount of time (this amount varies from an empirical investigation to another - generally between 1 min and 3 min). Then, the participants proceed on doing the same with the next item, etc.

The evaluation of the Alternative Uses Test is done on Fluency, Flexibility and Originality or Novelty.

[1] http://corpus.byu.edu/coca/

[2] http://ucrel.lancs.ac.uk/claws7tags.html

Description of the knowledge acquisition for the artificial cognitive system solver

An Object Replacement Object Composition (OROC) system was deemed able to give similar answers as humans to the Alternative Uses Test [Olteţeanu and Falomir, 2015b].

OROC had a knowledge base of 70 simple and 20 composed objects. These objects are described through their various features (name, material, shape, affordance, size). Their features were manually encoded from descriptions of the object, and considered common sense knowledge. OROC makes creative inferences of similarity of affordance based on similarity of other features - e.g. if object a has affordance aff_a, than objects b and c can be proposed as having the same affordances or being suitable replacements if they have similar (functional) properties.

Description of evaluation with human data

Five objects from a household items domain were chosen for OROC to deploy its creative inference of affordance on. These objects were: Cup, $Newspaper$, $Toothbrush$, $Carpet$, $Dental\ Floss$. Thus, for an object like Cup, OROC could come up with creative inferences about its affordances, e.g. *A cup can be used for putting flowers in* (based on its similarity of shape with known object $Vase$).

Fluency and flexibility of OROC's answers was assessed in the same manner as for the evaluation of human answers. Human judges were then employed to assess Novelty, Usefulness and Likability on a 1-7 scale. Human judges were not informed that they were assessing the creative answers of an artificial system. The most novel, useful and likable items were thus classified (for example, the alternative use response judged as demonstrating most novelty was *Dental floss may be used to hang clothes to dry*). After evaluation, the system demonstrated similar ratings as those obtained by answers given by human participants. As a side result, a correlation between ratings on Usefulness and Likability was observed in human judges. More importantly, this evaluation demonstrated the ability to apply similar techniques in assessing creative solving by human and artificial cognitive systems.

It is worth noting that OROC was not implemented with the Alternative Uses Test in mind, but for the general purpose of being able to replace object a with an object b with similar functionality when object a was in its KB, but not in the environment. Part of OROC's abilities, like that of composing new objects, remain presently untested via comparison to a human creativity task counterpart.

5 General Approach

As illustrated in the previous case studies, a general approach can be formulated for producing artificial cognitive systems that can yield comparable results to humans in creativity tests.

This approach involves several steps, listed in the following:

1. Choosing a human creativity test the results of which are to be replicated via a cognitive system, or choosing a creative problem-solving skill that is more general (like creative object replacement) and has some empirical adjacent validation possible (like the Alternative Uses test for object replacement as a general skill).
2. Finding a source of knowledge for cognitive knowledge acquisition. Implementing types of knowledge acquisition or organization which are cognitively inspired, or which might yield further cognitive results. For example, the OROC solver demonstrates other cognitive effects, like shape bias [Imai *et al.*, 1994].
3. Implementing a system which uses processes similar to (or is able to produce results similar to) cognitive creative processes, like association, re-representation, etc.
4. Using human normative data for that particular test or general task; obtaining normative data by running that creativity test if data is not readily available.
5. Evaluating the results of the artificial cognitive system using: a) human normative data and/or b) evaluation techniques used for assessing the human creativity task.
6. Deploying data analysis measures which enable new possible relations of scientific interest to be observed.
7. Enabling the artificial cognitive system with generative abilities for that particular test or task (if possible). This will allow for new empirical testing of human participants with controlled variables, to further refine hypotheses about creative processes.

6 Applicability of these principles to other creativity tests

In this section, the manner in which the steps of the general approach have been implemented in the two case studies is summarized. The steps that can still be implemented are explained. An analysis of the applicability of these steps to other creativity tests is performed.

6.1 The RAT solver

The Remote Associates Test [Mednick and Mednick, 1971] has been previously described (Section 4.1). The approach steps have been implemented by Olteţeanu and Falomir[Olteţeanu and Falomir, 2015a] as follows:

1. Human creativity test used: The Remote Associates Test.
2. Knowledge used: 2-grams from a human language corpus. Other language corpuses can also be used to check for scalability. Knowledge acquisition process: Use of a cognitive framework, Concepts, Links.
3. Cognitively inspired processes: Association based search, Convergence (as explained in [Olteţeanu and Falomir, 2015a]).
4. Normative data used: Normative data for 144 queries from Bowden and Jung-Beeman [2003].
5. Evaluation used: ability to respond to the same queries as the ones in human normative data; correlation between answer probability and human difficulty.
6. Data analysis measures and observations: correlation with human difficulty; yet to be empirically tested on a significant scale: influences of frequency on answer

preference. Possible other mesures: influences of frequency on appreciation of RAT query-answer pair as more creative.

7. Generative abilities: mentioned as available. Can offer control over frequency, presentation order of items, etc.

6.2 The Alternative Uses solver system

The Alternative Uses Test [Guilford, 1967] has been previously described (Section 4.2). The general approach can be illustrated stepwise for the implementation by Olteţeanu and Falomir [2015b] as follows:

1. Human creativity test used: The Alternative Uses Test.
2. Knowledge used: Object, Object parts and Features Knowledge. Knowledge organization process: Anchoring objects (concepts) in feature maps, ordering feature maps based on feature similarity.
3. Cognitively inspired processes: Similarity-based association, structure (transfer), shape bias.
4. Normative data used: Evaluation of human answers to the Alternative Uses Test by Gilhooly et al [2007].
5. Evaluation used: Fluency and Flexibility; Novelty, Usefulness and Likability as assessed by human judges. Comparability of these ratings to those of human responses. OROC's processes were analysed in comparison with the processes used by human participants as obtained via a think aloud protocol.
6. Data analysis measures and observations: Relation between Novelty, Usefulness and Likability. Other possible measures: analysis of different features as driving human responses via similarity - to be compared with such features in the system.
7. Generative abilities: possibly available. Could offer control over various types of feature influence in creative judgement, to answer questions such as: "Is shape more often used than material when making creative inferences?", "Can modifications of shape enable/disable the creative inference for humans?", etc.

6.3 The Wallach-Kogan test

The Wallach-Kogan test [1965] gives participants a specific property or component, and requires them to enumerate as many items as they can which have that property or contain that component (e.g. items that are green, that make noise, that have wheels, etc.). These are possible steps of building or making comparable an artificial cognitive system:

1. Human creativity test: the Wallach-Kogan test.
2. Knowledge that can be used: Object and Features Knowledge. Knowledge acquisition process: data from human descriptions of objects can be parsed for this purpose. Tasks can be given over crowdsourcing platforms to gather more such data.
3. Cognitively inspired processes: Similarity-based association, common parts/structure, feature based search.
4. Normative data can be acquired by giving the Wallach-Kogan test to human participants, with stimuli consisting of sets of oftenly encountered properties and object parts. Thus, normative data for flexibility, fluency and originality of answers can be gathered. Human judges can be used for novelty and elaboration ratings.
5. Evaluation: Fluency, Flexibility, Novelty, Originality.
6. Data analysis measures and observations: Relationships between feature access and component access speed, between speed of access and frequency of property (in human answers), between Fluency and Novelty ratings, etc.
7. Generative abilities: can be used to control for frequency of objects-properties or objects-components relations, thus helping investigate (i) whether feature based search and component based search of objects which have those features and properties are processes with equal or different speed in human participants; (ii) whether different types of features based queries have are easier or harder than others; (iii) whether number of components in the object yielded influences performance, etc.

6.4 Insight tests

Insight tests, like the ones by Maier [1931] and Duncker [1945] require a larger amount of knowledge and heuristics in order to be succesfully solved and implemented. Addressing insight tests which require object knowledge before those which require abstract knowledge might be a productive scalable strategy.

1. Human creativity test: Object insight tests.
2. Knowledge used: Object knowledge, problem templates and heuristics. Knowledge acquisition process: knowledge of object properties has already been discussed; knowledge of object affordances can be acquired in a similar way. Tasks can be given over crowdsourcing platforms to acquire affordances of sets of commonly used objects. Data mining strategies can be used to extract from text corpora sets of objects with the verbs that are used in conjunction with them - a subset of which will constitute the object's affordances. It is sensible to consider that at least a subset of the problem templates which pertain to objects will be constructed from such affordances.
3. Cognitively inspired processes: Re-representation, creation of new problem templates, creation of new heuristics.
4. Normative data: Data on object insight tests can be acquired from existing sources [Jacobs and Dominowski, 1981], or via new empirical investigations.
5. Evaluation tools that can be used: comparability to humans in think aloud protocols, the use/creation of similar problem templates, knowledge and processes, switch between templates, escape from functional fixedness.
6. Data analysis measures and observations: problem templates regularly employed, number of associated objects, their relation to response times, etc.

7. Generative abilities: control over variables like problem templates used or heuristics triggered, objects, features.

7 Cognitive knowledge acquisition from creativity tests

Some human creativity tests can be used to provide knowledge bases for artificial cognitive systems. For example, the Wallach-Kogan test will yield a set of data pertaining to object properties and object parts, as for each property or object part various objects having that property or that object part are asked for from the participant.

These answers can be used in the knowledge base of artificial systems, in the context of other tests - like the Alternative Uses test - where various properties are required to be known by the system, or the object insight tests, in which properties might be relevant for future affordance in solving object tasks.

Furthermore, common answers in the Wallach-Kogan test can be considered as high-frequency associates, and models can be built to interpret the frequency of occuring answers in the Wallach-Kogan test or their ordering as weigths of associative links.

Similarly, giving human participants a set of object-related insight problems using the think aloud protocol will provide knowledge on problem templates used (even if some of them are not productive for that particular problem), from which observations can be made on how such templates are constructed.

8 Discussion

As shown in the previous case studies (Sections 4.1 and 4.2), this approach is useful from both an Artificial Intelligence and a cognitive science perspective.

From the AI perspective, new techiques can be developed out of the inspiration of human creative cognitive processing.

From the cognitive science perspective, these systems can further be used by cognitive psychologists for the more detailed cognitive modeling of creative tasks, the implementation and testing of various theoretical hypotheses on how creative processes proceed and how knowledge organization sustains such processes.

New relations can be observed during the implementation of systems capable of comparable results in data analysis, like the correlation between the probability of finding an answer and the difficulty of query for the (RAT, comRAT) test–system pair, or the relationship between Usefulness and Likability in the Alternative Uses Test.

In order to bridge the gap between computational creativity and empirical research on human creativity, we must aim to address not only artefact creating systems, but also problem-solving and creative reasoning systems, and systems which can give answers comparable to humans in creativity tests. This will allow the further study of the creative process, in both artificial and natural cognitive systems.

In conclusion, this paper has presented a general approach to building cognitive computational creativity systems which can give comparable answers as humans to human creativity tests. Two case studies of systems which realize this were presented. A set of steps was laid down as possible methodology when approaching such systems. The application of these steps to other creativity tests was briefly explored.

Acknowledgments

The support by the Deutscher Akademischer Austauschdienst (DAAD) and by the Bremen Cognitive Systems and Spatial Cognition Centre at Universität Bremen is acknowledged.

References

[Ansburg, 2000] Pamela I Ansburg. Individual differences in problem solving via insight. *Current Psychology*, 19(2):143–146, 2000.

[Baba, 1982] Yuji Baba. An analysis of creativity by means of the Remote Associates Test for Adult Revised in Japanese (JARAT FORM A). *Japanese Journal of Psychology*, 1982.

[Batchelder and Alexander, 2012] William H. Batchelder and Gregory E. Alexander. Insight problem solving: A critical examination of the possibility of formal theory. *The Journal of Problem Solving*, 5(1):56–100, 2012.

[Boden, 2003] M.A. Boden. *The Creative Mind: Myths and Mechanisms*. Routledge, 2003.

[Bowden and Jung-Beeman, 2003] Edward M Bowden and Mark Jung-Beeman. Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*, 35(4):634–639, 2003.

[Chermahini *et al.*, 2012] Soghra Akbari Chermahini, Marian Hickendorff, and Bernhard Hommel. Development and validity of a dutch version of the remote associates task: An item-response theory approach. *Thinking Skills and Creativity*, 7(3):177–186, 2012.

[Colton *et al.*, 2011] Simon Colton, A Pease, and J Charnley. Computational creativity theory: The FACE and IDEA descriptive models. In *Proceedings of the Second International Conference on Computational Creativity*, pages 90–95, 2011.

[Colton *et al.*, 2012] Simon Colton, Jacob Goodwin, and Tony Veale. Full-FACE poetry generation. In *Proceedings of the Third International Conference on Computational Creativity*, pages 95–102, 2012.

[Colton, 2012] Simon Colton. The painting fool: Stories from building an automated painter. In *Computers and creativity*, pages 3–38. Springer, 2012.

[Dorfman *et al.*, 1996] Jennifer Dorfman, Victor A Shames, and John F Kihlstrom. Intuition, incubation, and insight: Implicit cognition in problem solving. *Implicit cognition*, pages 257–296, 1996.

[Duncker, 1945] K. Duncker. On problem solving. *Psychological Monographs*, 58(5, Whole No.270), 1945.

[Evans, 1964] Thomas G Evans. A heuristic program to solve geometric-analogy problems. In *Proceedings of the April 21-23, 1964, spring joint computer conference*, pages 327–338. ACM, 1964.

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

[Forbus *et al.*, 1995] Kenneth D Forbus, Dedre Gentner, and Keith Law. Mac/fac: A model of similarity-based retrieval. *Cognitive Science*, 19(2):141–205, 1995.

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

[Gilhooly *et al.*, 2007] KJ Gilhooly, E Fioratou, SH Anthony, and V Wynn. Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, 98(4):611–625, 2007.

[Guilford, 1967] Joy Paul Guilford. *The nature of human intelligence.* McGraw-Hill, 1967.

[Halford *et al.*, 1994] Graeme S Halford, William H Wilson, Jian Guo, Ross W Gayler, Janet Wiles, and JEM Stewart. Connectionist implications for processing capacity limitations in analogies. *Advances in connectionist and neural computation theory*, 2:363–415, 1994.

[Hamilton, 1982] Marlene A Hamilton. " Jamaicanizing" the Mednick Remote Associates Test of creativity. *Perceptual and Motor Skills*, 55(1):321–322, 1982.

[Hélie and Sun, 2010] Sébastien Hélie and Ron Sun. Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological review*, 117(3):994, 2010.

[Hofstadter *et al.*, 1994] Douglas R Hofstadter, Melanie Mitchell, et al. The Copycat project: A model of mental fluidity and analogy-making. *Advances in connectionist and neural computation theory*, 2(31-112):29–30, 1994.

[Hummel and Holyoak, 1997] John E Hummel and Keith J Holyoak. Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, 104(3):427, 1997.

[Imai *et al.*, 1994] Mutsumi Imai, Dedre Gentner, and Nobuko Uchida. Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development*, 9(1):45–75, 1994.

[Jacobs and Dominowski, 1981] Mary K Jacobs and Roger L Dominowski. Learning to solve insight problems. *Bulletin of the Psychonomic Society*, 17(4):171–174, 1981.

[Kim, 2006] Kyung Hee Kim. Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, 18(1):3–14, 2006.

[Lenat, 1976] Douglas B Lenat. Am: An artificial intelligence approach to discovery in mathematics as heuristic search. Technical report, DTIC Document, 1976.

[Maier, 1931] Norman RF Maier. Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, 12(2):181, 1931.

[Mednick and Mednick, 1971] Sarnoff A Mednick and Martha Mednick. *Remote associates test: Examiner's manual*. Houghton Mifflin, 1971.

[Nevo and Levin, 1978] B Nevo and I Levin. Remote associates test: assessment of creativity in Hebrew. *Megamot*, 24:87–98, 1978.

[Olteţeanu and Falomir, 2015a] Ana-Maria Olteţeanu and Zoe Falomir. comRAT-C: A computational compound remote associate test solver based on language data and its comparison to human performance. *Pattern Recognition Letters*, Special Issue: Cognitive Systems for Knowledge Discovery, 2015.

[Olteţeanu and Falomir, 2015b] Ana-Maria Olteţeanu and Zoe Falomir. Object replacement and object composition in a creative cognitive system. towards a computational solver of the Alternative Uses Test. *Cognitively Inspired AI Systems*, submitted, 2015.

[Olteţeanu, 2014] Ana-Maria Olteţeanu. Two general classes in creative problem-solving? An account based on the cognitive processes involved in the problem structure - representation structure relationship. In T. Besold, K.-U. Kühnberger, M. Schorlemmer, and A. Smaill, editors, *Proceedings of the Workshop Computational Creativity, Concept Invention, and General Intelligence*, volume 01-2014 of *Publications of the Institute of Cognitive Science*, Osnabrück, 2014. Institute of Cognitive Science.

[Olteţeanu and Falomir, 2014] A.-M. Olteţeanu and Z. Falomir. Towards a Remote Associate Test Solver based on Language Data. In Lledo Museros et al., editor, *Artificial Intelligence Research and Development*, volume 269 of *Frontiers in Artificial Intelligence and Applications*, pages 249–252. IOS Press, 2014.

[Pease *et al.*, 2001] Alison Pease, Daniel Winterstein, and Simon Colton. Evaluating machine creativity. In *Workshop on Creative Systems, 4th International Conference on Case Based Reasoning*, pages 129–137, 2001.

[Ritchie, 2001] Graeme Ritchie. Assessing creativity. In *Proc. of AISB'01 Symposium*. Citeseer, 2001.

[Wallach and Kogan, 1965] Michael A Wallach and Nathan Kogan. *Modes of thinking in young children: A study of the creativity-intelligence distinction.* Holt, Rinehart & Winston, 1965.

[Wiggins, 2001] Geraint A Wiggins. Towards a more precise characterisation of creativity in AI. In *Case-based reasoning: Papers from the workshop programme at ICCBR*, volume 1, pages 113–120, 2001.

[Williams and McOwan, 2014] Howard Williams and Peter W McOwan. Magic in the machine: a computational magician's assistant. *Frontiers in psychology*, 5, 2014.

[Worthen and Clark, 1971] Blaine R Worthen and Philip M Clark. Toward an improved measure of remote associational ability. *Journal of Educational Measurement*, 8(2):113–123, 1971.