

# Artificial Cognitive Systems that can answer Human Creativity Tests: An Approach and Two Case Studies

Ana-Maria Oltețeanu, Zoe Falomir, Christian Freksa  
Bremen Spatial Cognition Center, Universität Bremen

**Abstract**—Creative cognitive systems are rarely assessed with the same tools as human creativity. In this paper, an approach is proposed for building cognitive systems which can solve human creativity tests. The importance of using cognitively viable processes, cognitive knowledge acquisition and organization, and cognitively comparable 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.

**Index Terms**—Computational Creativity, Human Creativity, Cognitive Systems, Cognitive Processing, Cognitive Knowledge Acquisition, Cognitively-comparable Evaluation

## I. INTRODUCTION

WE envision a future where cognitive agents help humans solve problems in their daily routines at home and at work. In order to be helpful, those agents must find good solutions or creative alternatives the human can select from. Imagine you are in your house, needing to solve a particular problem. You need a particular tool or object (e.g. a piece of string) or recipe ingredient (e.g. mince meat) but you do not 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 imagine you are in a situation in which you ran out of ideas on how to solve a particular problem, and would like to think of a completely different approach. Let us say that you would like the help of a system that could inspire you. Such a system would need to operate or at least communicate in a cognitive manner – understand what type of information is associated with the problem at hand, in a manner that would enhance your creative process – for example by predisposing you to new ways of seeing the problem or re-representing its content. It could show you items (websites, research articles, excerpts from encyclopedias, music, films, photos, formulas) which could trigger for you new associations, new ways of solving or framing that problem.

Corresponding author: Ana-Maria Oltețeanu, amoodu@informatik.uni-bremen.de

Manuscript received...; revised...

Furthermore, 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 [1], music [2], painting [3] to mathematics [4]. Empirical tests of human creativity and creative problem-solving do exist [5], [6], [7], [8], [9], [10]. 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. [11]), and thus not as much progress is made as possible 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 that can be assessed with human creativity tests. To support this claim, this paper presents a general approach for building artificial cognitive systems that can solve human creativity tests in a cognitive manner. The types of knowledge, knowledge acquisition, cognitive processes and cognitive evaluation which can be used are also discussed. 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. Two case studies of systems which can solve human creativity tests are then presented, briefly describing how cognitive knowledge was acquired and organized for such systems, the processes used and how the systems were evaluated compared to their human counterparts.

The rest of this paper is structured as follows. Section II 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 III. Sections IV-A and IV-B briefly describe two case studies of artificial cognitive systems which can give comparable results to humans in cre-

ativity tests, together with the knowledge acquisition, cognitive processing and evaluation of these systems. Section V presents a general approach towards making artificial cognitive systems that yield comparable results to humans in creativity tests. The applicability of this approach to other creativity tests is discussed in Section VI.

A short description of how cognitive knowledge acquisition can be performed using creativity tests is provided in section VII. A short discussion and conclusion is presented in Section VIII.

## II. 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 [12], [13]. Some authors aim to produce artefacts through their systems, and would like them to be assessed as having comparable creativity to that of humans. Others define creativity in terms of process [14], not necessarily aiming for an implementation. Ritchie [13] proposed an assessment of computational creative systems which takes into account the inspiring set – a union of implicit and explicit knowledge – formalizing 14 criteria of evaluation, centered around typicality, quality and novelty (see also [15]). Pease, Winterstein and Colton [16] 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 [17] 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 [17] uses the notion of the 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 [18]), asking users to rate the products of computational creativity systems, or to choose words which appropriately describe their reaction to those products. Yet other thoughts on evaluation can be seen in the move against mere generation commented upon by Ventura [19], Jordanous' SPECS methodology [20], the discussion on creativity versus the perception of creativity and Colton's creative tripod [21], etc.

Although the field of computational creativity evaluation has made significant progress, none of the types of evaluation above shines a bright light on cognitive creative processing. Some steps in the direction of human cognition are being made, with some research addressing human tasks [22], and models being based on a psychological theory of communicative interaction [23].

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

- The Remote Associates Test [5];

- The Alternative Uses Test [6];
- The Torrance Creativity Tests (reviewed by Kim [7]);
- The Wallach-Kogan tests [8];
- Insight tests (like the ones described by [9], [10])

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) Success: 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 [6], the Wallach-Kogan test [8]). 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.

## III. 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 [24], MAC/FAC [25], LISA [26], STAR [27], Copycat [28], etc.). Another creative process which has recently gathered investigative interest is that of *conceptual blending* [29]. Concept blending is defined by Fauconnier [29] as a basic mental operation of constructing a partial match between two inputs, projecting selectively from those inputs into a “blended space” which has as a result the emergence of a new structure and of new meaning. The COINVENT project investigated and computationally instantiated this process [30]. Various theories account for *insight problem-solving* [31]: for example, this type of problem solving is assumed to include four stages – Preparation, Incubation, Illumination or Insight and Verification or Evaluation. An important process which is considered part of insightful problem solving is that of re-representation. 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 creativity tasks. In the authors' opinion [32], 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**, as follows. 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 (like  $c$  and/or  $d$ ) 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 [33], replacement of structure parts [34], navigation between similar structures, and structure-based operations (merging, overlap, removal of unnecessary parts, etc.) [32]. These allow 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 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 of giving results comparable to humans in various creativity tests.

#### IV. CASE STUDIES

This section summarizes two case studies of cognitive systems which yield results comparable to human participants in creativity tests. Section IV-A presents comRAT, a cognitive system which can give comparable answers to humans in the Remote Associates Test [35]. Section IV-B presents OROC, a prototype system capable of giving similar answers to humans to the Alternative Uses Test [34].

##### A. Case 1 - comRAT - A Remote Associates Test solver

*The Remote Associates Test:* The Remote Associates Test (RAT) by Mednick and Mednick [5] is a creativity test in which participants are given three word items and required to produce a fourth, which is associated 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 [36] - structural or compound items reflect language based relationships, and functional items reflect relationships beyond language.

The RAT has been widely used as an empirical measurement of creativity [37], [38] and translated<sup>1</sup> into multiple languages

<sup>1</sup>Translation of the RAT is not a trivial task, as a RAT item in English will rarely translate in a RAT item in Italian, for example, as associates between the query words and the answer might not exist, or be reasonable compounds. In that context, a translation of the RAT generally means creating a set of suitable items for the test in that language, and checking for the validity of those items. The compound RAT thus depends on language and language use, whereas the functional RAT reflects semantic relations within a language.

[39], [40], [41], [42].

The type of data acquired after empirical testing includes the percentage of participants solving a particular test item and response speed. Useful sources of normative data are available [43].

*Description of the knowledge acquisition and solving process for the artificial cognitive system solver:* A computational RAT solver (comRAT) was presented by Oltejeanu and Falomir [44], [35], using language data and a convergence process, which will further be briefly summarized. The language data used consisted of the 1,048,720 most frequent 2-grams from the Corpus of Contemporary American English<sup>2</sup>. The University of Lancaster – Constituent Likelihood Automatic Word-tagging System (UCREL CLAWS7) tagset<sup>3</sup> provides grammatical tags to words in bodies of continuous text. Based on this tagset, 205,602 2-grams with relevant tags for the task were kept (a complete list of the types of words kept based on their tag is provided by [35]). 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* associatively attached 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 through an associative process.

*Description of evaluation with human data:* The comRAT system was given the 144 items from the Bowden and Jung-Beeman [43] normative data to solve, thus being tested with the exact same queries as humans. 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. From a cognitive point of view, computational systems can be used to model the human process and obtain similar performance, rather than solve tasks to perfection. When solving this task, comRAT is limited in its ability to come up with an answer by its existing knowledge, as human participants would also be. In 26 cases, comRAT comes up with other plausible answers – for example it answers the query MILL, TOOTH, DUST, for which the correct answer provided by the normative data is SAW, with the plausible answer GOLD; the query CRY, FRONT, SHIP, for which the correct answer from the normative data is BATTLE, is given the plausible answer WAR, etc.

Furthermore, using data on the frequency of 2-grams, the probability of the system to find an answer has been found to correlate with the difficulty of the RAT queries for humans. The probability to find an answer given each query word is calculated as the ratio of favourable cases in which the query word and answer word appear linked together, over the total number of cases in which the query word appears linked to all *Concepts*. The initial calculation considered each query word

<sup>2</sup><http://corpus.byu.edu/coca/>

<sup>3</sup><http://ucrel.lancs.ac.uk/claws7tags.html>

to have an equal influence in determining the probability of the answer, however this can be further modified to account for possible order effects. 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, as follows. The correlation between accuracy and comRAT's probability was  $r = 0.49, p < 0.002$  for mean solution time in 30s - thus the higher the probability, the more accurate people were on average in answering the query. The inverse correlation between response times and comRAT's probability was  $r = -0.52, p < 0.001$  for mean solution time in 30s - thus the higher the probability, the less time people took on average to solve the query.

This shows that the process of association and convergence used by comRAT to solve the task has high chances of providing further insights into future cognitive models of the task. As the system can come up with plausible answers even when lacking the knowledge to respond correctly, this demonstrates possible further uses of associative processes in order to ensure flexibility and robustness in knowledge-based systems. The system further showed an ability to give multiple answers to some queries - for example, to query HIGH, DISTRICT, HOUSE, it could provide answers SCHOOL, COURT, OFFICE, WATER and others. Answer SCHOOL (which is also the only answer provided as correct in Bowden and Jung-Beeman's normative dataset) "won" via its higher probability in comRAT. This capacity for multiple answers based on knowledge acquired from human data will further allow the cognitive empirical study of the reasons why some answers are preferred by humans over others, and the links between this and frequency of known expressions.

The comRAT solver can be used to generate new RAT queries. RAT items can thus be generated for future empirical work with human participants, controlling for semantic tag, position in the 2-gram of the given and answer word, frequency of the various query items and probability of finding the answer. This allows the generation of sets of RAT queries in which different variables are kept constant, thus making comRAT a future useful tool for the cognitive modeling community.

The comRAT solver and its principles have also been used to translate the RAT task to the visual domain. Though a validation of these items is yet in the making, comRAT could thus help examine creativity tests in a multimodal perspective [45].

The cognitive process of association and the cognitive data on frequency of two grams and other types of associates are thus used to create an artificial cognitive system which has many future possible applications in the cognitive modeling community. Using comRAT's ability to generate and solve queries will provide more knowledge about how the RAT is solved by humans, which can help improve comRAT further, and possibly tackle a wider array of association-based tasks in the future.

### B. Case 2 - OROC - an object replacement, object composition system which can solve the Alternative Uses Test

*The Alternative Uses Test:* The Alternative Uses Test [6] 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.

Normative data for the Alternative Uses Test can be easily obtained in relation to a set of objects, for example objects from the household domain. Such data should include human Fluency, Flexibility and Originality or Novelty, as well as order of responses - which might later be useful in understanding the types of creative processes employed to generate the alternative uses.

Evaluations on the same types of metrics as for human answers can be made for the artificial cognitive system (on Fluency, Flexibility and Originality or Novelty). Furthermore, the processes employed by the system can be compared to those employed by humans, as observed in think aloud protocols [46]. If normative order of responses has been obtained, then an analysis can be performed on the influence various object features have over coming up with new uses.

*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 [34].

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 in a cognitively-inspired process - determining 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 were selected from a household items domain on which OROC was to deploy its creative inference of affordance. 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*).

The fluency and flexibility of OROC's answers to these five objects - a total of 30 alternative object use statements - was assessed in the same manner in which the answer would have been assessed if provided by humans which were undergoing the Alternative Uses Test. Human judges were then employed to assess Novelty, Usefulness and Likability on a 1-7 Likert scale, where 1 represented the lower bound, for example 'Not at all novel', and 7 the upper bound, e.g. 'Highly novel'. Usefulness and Likability were new metrics added in this assessment besides the normally used Novelty metric. These

were meant to split the feedback from the judges in different dimensions, and thus dispel possible influences between how much a judge liked a particular use and how much they found it to be useful or novel.

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. The answers provided by OROC were rated as follows: mean Novelty 3.79 ( $SD = 1.69$ ), mean Likability 3.31 ( $SD = 1.68$ ), mean Usefulness 3.77 ( $SD = 1.7$ ). Human answers from Gilhooly et al. [46] are rated only for Novelty: 2.54 ( $SD = 0.74$ ) in a think aloud group and 2.45 ( $SD = 0.51$ ) in a silent group respectively. As a side result, a correlation between ratings on Usefulness and Likability was observed in human judges:  $r = 0.63$ ,  $p < 0.005$ . Note that, when asked to provide their ratings on Novelty, Usefulness and Likability, the judges were not provided with specific definitions, thus they were guided by their own subjective definition of these dimensions, making this evaluation procedure similar to the consensual assessment technique [47], in which independent individuals make subjective judgements on products in domains they are familiar with. As definitions of these dimensions are not a priori provided, creativity assessments are more likely to mirror assessments in the real world, a more accurate measure of interjudge reliability is obtained and the definitions of the experimenter are not imposed onto the judges.

OROC's *processes* of proposing new object affordances were also assessed and showed comparability to the cognitive processes deployed by humans when solving the Alternative Uses Test, as explored by a think aloud protocol by [46]. More importantly, these forms of evaluation demonstrated (i) the ability to apply similar techniques in assessing creative solving of human and artificial cognitive systems, and (ii) comparability of process which can be used for further empirical and computational investigations.

It is worth noting that OROC was not implemented with the Alternative Uses Test in mind, but for testing a mechanism from a cognitive creative problem solving theoretical framework [48], [32]. This mechanism had the general purpose of being able to replace object *a* with an object *b* with similar functionality, for cases when object *a* was in its knowledge base, 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.

## V. GENERAL APPROACH

A general approach can be formulated, based on the above examples, for producing artificial cognitive systems that can yield comparable results to humans in creativity tests, and aim for cognitive or cognitively-inspired processes and cognitive knowledge acquisition. This approach involves several steps:

- 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 and has some empirical adjacent validation possible.

- 2) Finding a source of knowledge for cognitive knowledge acquisition, or gathering such knowledge from human participants.
- 3) Implementing a system which uses processes similar to (or is able to produce results similar to) cognitive creative processes, like association, re-representation, use of context, structure and similarity, etc. Implementing types of knowledge organization which are cognitively inspired, or which might yield further cognitive results.
- 4) Finding or obtaining human normative data for that particular test or general task.
- 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. Supplementary, computational creativity evaluation procedures can also be applied.
- 6) Deploying data analysis measures which enable new possible relations of scientific interest to be observed. Such measures can be deployed on computational data, on empirical data and on the computational-empirical comparison.
- 7) Enabling the artificial cognitive system with generative abilities for that particular test or task (if possible) - i.e. with the possibility of not only solving a particular creative or creative problem-solving task, but also with the ability of creating more tasks of this type. This will allow for new empirical testing of human participants with controlled variables, to further refine hypotheses about creative processes.

This approach and its essential steps are shown in Fig.1.

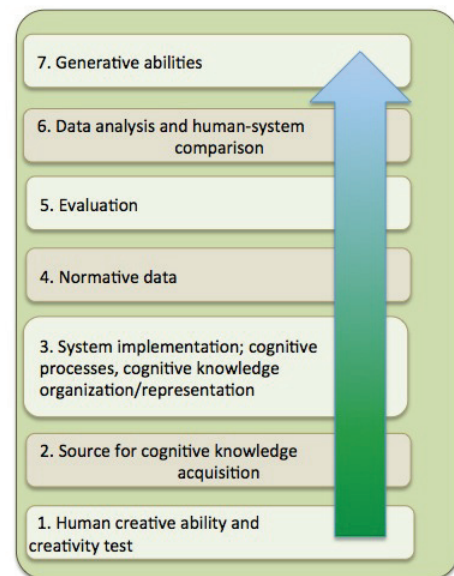


Fig. 1. General approach steps

## VI. APPLICABILITY OF THESE PRINCIPLES TO OTHER CREATIVITY TESTS

In this section, the Wallach-Kogan instances test [8] and insight tests [9], [10] are proposed as two other possible creativity tests on which this approach could be applied. Sections VI-A and VI-B describe these tests and particularities of applying this approach to implementing artificial cognitive systems for each of them. Table I gives an overview of the application of the general approach to the previously mentioned case studies (IV-A, IV-B), and to the other two proposed creativity tests (VI-A, VI-B) and corresponding skills.

### A. *The Wallach-Kogan instances test*

The Wallach-Kogan instances test [8] 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.).

As shown in the fourth column of Table I, 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. Cognitive knowledge acquisition can be performed by either a) asking human participants to describe objects in terms of their defining properties and components, or b) by machine learning approaches directed at acquiring such property-object or component-object co-occurrences from text corpora.

A computational cognitive agent that solves the Wallach-Kogan instances test could also be developed to have generative abilities, thus to create queries based on controlling for the frequency of relations between properties (or components) and objects. Giving human participants queries in which the frequency of object-properties and object-components relation is controlled would help us investigate (i) whether the speed of processes like property based search and component based search of objects is the same or not; (ii) whether different types of features-based queries are easier or harder than others; (iii) whether the number of components in the object influences performance, etc. This in turn will help us build more cognitively-informed creative problem-solving agents.

### B. *Insight tests*

Insight tests, like the ones by Maier [9] and Duncker [10] require a larger amount of knowledge and heuristics in order to be successfully solved and implemented. Addressing insight tests that require object knowledge before those which require abstract knowledge might be a productive scalable strategy. Such a strategy could for example employ data on objects and features that has been obtained for the Alternative Uses Test, or the Wallach-Kogan test.

Object affordances are required for solving practical insight tests as well. 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, paired 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 that pertain to objects will be constructed from such affordances. This knowledge can be complemented with the acquisition of object-related problem templates and common heuristics employed by human participants when approaching practical object insight tests. Such acquisition might require the analysis of think-aloud protocol data of human participants solving insight problems.

Normative data on object insight tests can be acquired from existing sources [49], or via meta-analysis on existing results on different tests. Such meta-analysis might be necessary because insight tests generally require a large amount of time to be deployed, thus the empirical data from one source is most often restricted to only a few insight tests, or even to one.

Among the multiple comparison and evaluation tools that can be employed, one of the most interesting ones is that of comparing the use, switch between, and creation of new problem-templates. An artificial cognitive agent performing such tasks of re-representation would undoubtedly prove of much interest for advancing the state of the art on creative reasoning techniques.

Generative abilities in a system which can solve insight problems, even in a restricted domain, would provide a reliable way to create and modify a larger set of insight problems.

## VII. 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 affordances 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 occurring answers in the Wallach-Kogan test or their ordering as weights 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.

## VIII. DISCUSSION

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

TABLE I  
SHOWING APPLICATION OF THE COMPARABILITY APPROACH TO VARIOUS CREATIVITY TESTS AND TASKS

Approach	A. RAT solver	B. AUT solver	C. Wallach-Kogan solver	D. Insight tests solver
1. Human creativity test or ability	The Remote Associates Test [5]	The Alternative Uses test [6]	Wallach-Kogan instances test [8]	Insights tests[9], [10]
2. Normative data	144 queries from Bowden and Jung-Beeman [43]	can be obtained for a pre-defined set of objects	can be obtained for a pre-defined set of properties and/or components	can be obtained by meta-analysis of insight tests papers or by new empirical experimentation
3. Source for cognitive knowledge acquisition	2-grams from language corpuses, word associates, ontologies	requires object, object parts and feature parts knowledge	requires object and features knowledge, affordance knowledge	object knowledge, problem templates and heuristics knowledge
4. System implementation; cognitive processes, cognitive knowledge organization/representation	association based search, convergence (as explained in [35])	affordances, similarity-based association, structure (transfer), shape bias	similarity-based association, common parts/structure, feature based search	re-representation, creation of new problem templates, creation of new heuristics.
5. Evaluation	ability to answer the same queries as humans, comparability to human normative data	comparison to human answers to the Alternative Uses Test obtained by Gilhooly et al [46], Fluency, Flexibility, Novelty, Originality metrics	Fluency, Flexibility, Novelty, Originality metrics; comparison to order of answers as produced by humans	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 and human-system comparison	relation to human response times, query difficulty, answer preference	analysis of different features as driving human responses via similarity - to be compared with such features in the system, the Novelty - Usefulness relation between alternative uses answers	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.	problem templates regularly employed, number of associated objects, their relation to response times, etc.
7. Generative abilities	control over frequency, item presentation order	control over various types of feature influence in producing alternative uses	can be used to control for frequency of objects-properties or objects-components relations	control over variables like problem templates used or heuristics triggered, objects, features

From the AI perspective, new techniques 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 unfold 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. Steps towards bridging this gap should involve (i) the use of cognitive processes and types of knowledge organization as an inspiration to some computational creativity systems; (ii) the use of computational creativity systems which have been tightly inspired by cognitive processes as potential models or tools to investigate human creativity processes; (iii) the use of cognitive knowledge acquisition techniques to inform the knowledge bases of such

systems; (iv) making results of some computational creativity systems and human creativity tasks directly comparable; this includes both comparing the results of systems (which can act as models) to human normative data, and the evaluation of computational creativity answers through techniques reserved for human participants; (v) the use of cognitive computational systems and models to generate rich new sets of controlled queries, which allow a deeper investigation of the human creative problem solving process.

In conclusion, this paper has presented a general approach to building cognitive computational creativity systems which can give comparable answers to humans when solving human creativity tests. A set of steps was laid down as possible methodology when approaching such systems. The application of these steps to various creativity tests was briefly explored and two case studies of systems which realize this approach were presented. As further work we plan to (i) implement and test this approach on the new cases described, and (ii) explore the applications of this approach to assistive systems - thus obtaining a deeper understanding of how natural cognitive systems can benefit and learn from the input of artificial creative cognitive systems.

#### ACKNOWLEDGMENTS

Ana-Maria Oltețeanu acknowledges the support of the Deutsche Forschungsgemeinschaft (DFG) for the *Creative Cognitive Systems*<sup>4</sup> (CreaCogs) project. Zoe Falomir acknowl-

<sup>4</sup>CreaCogs: <http://creacogcomp.com/>

edges the project *Cognitive Qualitative Descriptions and Applications*<sup>5</sup> (CogQDA), funded by the Universität Bremen through the *04-Independent Projects for Postdocs action*.

## REFERENCES

- [1] S. Colton, J. Goodwin, and T. Veale, "Full-FACE poetry generation," in *Proceedings of the Third International Conference on Computational Creativity*, 2012, pp. 95–102.
- [2] P. M. Todd and E. R. Miranda, "Putting some (artificial) life into models of musical creativity," in *Musical creativity: multidisciplinary research in theory and practice*, I. Deliège, G. A. Wiggins et al., Eds. Psychology Press, 2006, pp. 376–396.
- [3] S. Colton, "The painting fool: Stories from building an automated painter," in *Computers and creativity*. Springer, 2012, pp. 3–38.
- [4] D. B. Lenat, "AM: an artificial intelligence approach to discovery in mathematics as heuristic search," DTIC Document, Tech. Rep., 1976.
- [5] S. A. Mednick and M. Mednick, *Remote associates test: Examiner's manual*. Houghton Mifflin, 1971.
- [6] J. P. Guilford, *The nature of human intelligence*. McGraw-Hill, 1967.
- [7] K. H. Kim, "Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)," *Creativity Research Journal*, vol. 18, no. 1, pp. 3–14, 2006.
- [8] M. A. Wallach and N. Kogan, *Modes of thinking in young children: A study of the creativity-intelligence distinction*. Holt, Rinehart & Winston, 1965.
- [9] N. R. Maier, "Reasoning in humans. II. The solution of a problem and its appearance in consciousness," *Journal of Comparative Psychology*, vol. 12, no. 2, p. 181, 1931.
- [10] K. Duncker, "On problem solving," *Psychological Monographs*, vol. 58, no. 5, Whole No.270, 1945.
- [11] S. Hélie and R. Sun, "Incubation, insight, and creative problem solving: A unified theory and a connectionist model," *Psychological review*, vol. 117, no. 3, p. 994, 2010.
- [12] G. A. Wiggins, "Towards a more precise characterisation of creativity in AI," in *Case-based reasoning: Papers from the workshop programme at ICCBR*, vol. 1, 2001, pp. 113–120.
- [13] G. Ritchie, "Assessing creativity," in *Proc. of AISB'01 Symposium*. Citeseer, 2001.
- [14] M. Boden, *The Creative Mind: Myths and Mechanisms*. Routledge, 2003.
- [15] G. Ritchie, "Some empirical criteria for attributing creativity to a computer program," *Minds and Machines*, vol. 17, no. 1, pp. 67–99, 2007.
- [16] A. Pease, D. Winterstein, and S. Colton, "Evaluating machine creativity," in *Workshop on Creative Systems, 4th International Conference on Case Based Reasoning*, 2001, pp. 129–137.
- [17] S. 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*, 2011, pp. 90–95.
- [18] H. Williams and P. W. McOwan, "Magic in the machine: a computational magician's assistant," *Frontiers in psychology*, vol. 5, 2014.
- [19] D. Ventura, "Mere generation: Essential barometer or dated concept?" in *Proceedings of the Seventh International Conference on Computational Creativity*, 2016.
- [20] A. Jordanous, "A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative," *Cognitive Computation*, vol. 4, no. 3, pp. 246–279, 2012.
- [21] S. Colton, "Creativity versus the perception of creativity in computational systems," in *AAAI spring symposium: creative intelligent systems*, 2008, pp. 14–20.
- [22] D. A. Joyner, D. Bedwell, C. Graham, W. Lemmon, O. Martínez, and A. K. Goel, "Using human computation to acquire novel methods for addressing visual analogy problems on intelligence tests," in *Proceedings of the Sixth International Conference on Computational Creativity June*, 2015, p. 23.
- [23] E. R. Miranda, "Mimetic development of intonation," in *Music and Artificial Intelligence*. Springer, 2002, pp. 107–118.
- [24] T. G. Evans, "A heuristic program to solve geometric-analogy problems," in *Proceedings of the April 21-23, 1964, spring joint computer conference*. ACM, 1964, pp. 327–338.
- [25] K. D. Forbus, D. Gentner, and K. Law, "MAC/FAC: A model of similarity-based retrieval," *Cognitive Science*, vol. 19, no. 2, pp. 141–205, 1995.
- [26] J. E. Hummel and K. J. Holyoak, "Distributed representations of structure: A theory of analogical access and mapping," *Psychological Review*, vol. 104, no. 3, p. 427, 1997.
- [27] G. S. Halford, W. H. Wilson, J. Guo, R. W. Gayler, J. Wiles, and J. Stewart, "Connectionist implications for processing capacity limitations in analogies," *Advances in connectionist and neural computation theory*, vol. 2, pp. 363–415, 1994.
- [28] D. R. Hofstadter, M. Mitchell et al., "The Copycat project: A model of mental fluidity and analogy-making," *Advances in connectionist and neural computation theory*, vol. 2, no. 31-112, pp. 29–30, 1994.
- [29] G. Fauconnier and M. Turner, "Conceptual integration networks," *Cognitive science*, vol. 22, no. 2, pp. 133–187, 1998.
- [30] M. Schorlemmer, A. Smaill, K.-U. Kühnberger, O. Kutz, S. Colton, E. Cambouropoulos, and A. Pease, "Coinvent: Towards a computational concept invention theory," 2014.
- [31] W. H. Batchelder and G. E. Alexander, "Insight problem solving: A critical examination of the possibility of formal theory," *The Journal of Problem Solving*, vol. 5, no. 1, pp. 56–100, 2012.
- [32] A.-M. Oltețeanu, "Two general classes in creative problem-solving? An account based on the cognitive processes involved in the problem structure - representation structure relationship," in *Proceedings of the Workshop Computational Creativity, Concept Invention, and General Intelligence*, ser. Publications of the Institute of Cognitive Science, vol. 01-2014, Osnabrück, 2014.
- [33] D. Gentner, "Structure-mapping: A theoretical framework for analogy," *Cognitive science*, vol. 7, no. 2, pp. 155–170, 1983.
- [34] A.-M. Oltețeanu and Z. Falomir, "Object replacement and object composition in a creative cognitive system. towards a computational solver of the Alternative Uses Test," *Cognitive Systems Research*, 2016.
- [35] —, "comRAT-C: A computational compound remote associate test solver based on language data and its comparison to human performance," *Pattern Recognition Letters*, vol. 67, pp. 81–90, 2015.
- [36] B. R. Worthen and P. M. Clark, "Toward an improved measure of remote associational ability," *Journal of Educational Measurement*, vol. 8, no. 2, pp. 113–123, 1971.
- [37] P. I. Ansburg, "Individual differences in problem solving via insight," *Current Psychology*, vol. 19, no. 2, pp. 143–146, 2000.
- [38] J. Dorfman, V. A. Shames, and J. F. Kihlstrom, "Intuition, incubation, and insight: Implicit cognition in problem solving," *Implicit cognition*, pp. 257–296, 1996.
- [39] Y. 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.
- [40] M. A. Hamilton, "Jamaicanizing the Mednick Remote Associates Test of creativity," *Perceptual and Motor Skills*, vol. 55, no. 1, pp. 321–322, 1982.
- [41] B. Nevo and I. Levin, "Remote associates test: assessment of creativity in Hebrew," *Megamot*, vol. 24, pp. 87–98, 1978.
- [42] S. A. Chermahini, M. Hickendorff, and B. Hommel, "Development and validity of a Dutch version of the remote associates task: An item-response theory approach," *Thinking Skills and Creativity*, vol. 7, no. 3, pp. 177–186, 2012.
- [43] E. M. Bowden and M. Jung-Beeman, "Normative data for 144 compound remote associate problems," *Behavior Research Methods, Instruments, & Computers*, vol. 35, no. 4, pp. 634–639, 2003.
- [44] A.-M. Oltețeanu and Z. Falomir, "Towards a Remote Associate Test Solver based on Language Data," in *Artificial Intelligence Research and Development*, ser. Frontiers in Artificial Intelligence and Applications, L. M. et al., Ed. IOS Press, 2014, vol. 269, pp. 249–252.
- [45] A.-M. Oltețeanu, B. Gautam, and Z. Falomir, "Towards a visual remote associates test and its computational solver," in *Proceedings of the Third International Workshop on Artificial Intelligence and Cognition 2015*, vol. 1510. CEUR-Ws, 2015, pp. 19–28.
- [46] K. Gilhooly, E. Fioratou, S. Anthony, and V. Wynn, "Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects," *British Journal of Psychology*, vol. 98, no. 4, pp. 611–625, 2007.
- [47] B. A. Hennessey and T. M. Amabile, "Consensual assessment," *Encyclopedia of creativity*, vol. 1, pp. 347–359, 1999.
- [48] A.-M. Oltețeanu, "From simple machines to Eureka in four not-so-easy steps. Towards creative visuospatial intelligence," in *Fundamental Issues of Artificial Intelligence*, ser. Synthese Library, V. Müller, Ed. Springer, 2016, vol. 376, pp. 159–180.
- [49] M. K. Jacobs and R. L. Dominowski, "Learning to solve insight problems," *Bulletin of the Psychonomic Society*, vol. 17, no. 4, pp. 171–174, 1981.

<sup>5</sup>CogQDA:<https://sites.google.com/site/cogqda/>