CONTENTS

The logo was designed by using the social interaction and collaboration of ants foraging (in Stochastic Diffusion Search) and birds flocking (in Particle Swarm Optimisation).

The AISB16's 3rd International Symposium on Computational Creativity (CC2016)

Over the last few decades, computational creativity has attracted an increasing number of researchers from both arts and science backgrounds. Philosophers, cognitive psychologists, computer scientists and artists have all contributed to and enriched the literature.

Many argue a machine is creative if it simulates or replicates human creativity (e.g. evaluation of AI systems via a Turing-style test), while others have conceived of computational creativity as an inherently different discipline, where computer generated (art)work should not be judged on the same terms, i.e. as being necessarily producible by a human artist, or having similar attributes, etc.

This symposium aims to bring together researchers to discuss recent technical and philosophical developments in the field, and the impact of this research on the future of our relationship with computers and the way we perceive them: at the individual level where we interact with the machines, the social level where we interact with each other via computers, or even with machines interacting with each other.

The work included in this, the third annual installment of the AISB International Symposium on Computational Creativity, represents the continuing expansion of the field in ever more multidisciplinary directions. In terms of generative systems, this year's symposium includes talks on music generation, swarm intelligence based artwork, and narrative modelling. Additionally, papers considering questions about the evaluation of computational creativity and the place of more general issues of problem solving in the field are presented.

The theme that continues to emerge from progress in this area is one of inclusiveness and topical diversity: as the field matures, the scope of activities that are addressed with computationally creative systems broadens. At the same time, technical work on creative systems takes on an increasingly complex character, with researchers increasingly augmenting traditional symbolic and heuristic AI approaches for artefact generation with insights from parallel research in machine learning and dynamic systems.

We would like to thank all the members of the Programme Committee for the generous support and excellent work in evaluating the submissions.

Mohammad Majid al-Rifaie, Goldsmiths, University of London, UK
Stephen McGregor, Queen Mary University of London, UK

Organising Committee

- Mohammad Majid al-Rifaie (Goldsmiths, University of London, UK)
- Stephen McGregor (Queen Mary University of London, UK)
- Mohammad Ali Javaheri (Goldsmiths, University of London, UK) - publicity

Programme Committee

- Tarek Besold (Free University of Bozen-Bolzano, Italy)
- Mark Bishop (Goldsmiths, University of London, UK)
- Paul Brown (University of Sussex, UK)
- Amílcar Cardoso (University of Coimbra, Portugal)
- Simon Colton (Falmouth University, UK)
- Pablo Gervás (Universidad Complutense de Madrid, Spain)
- Bipin Indurkhya (AGH University of Science and Technology, Kraków, Poland)
- Colin Johnson (University of Kent, UK)
- Anna Jordanous (University of Kent, UK)
- Penousal Machado (University of Coimbra, Portugal)
- Jon McCormack (Monash University, Australia)
- François Pachet (SONY Computer Science Laboratory Paris, France)
- Alison Pease (University of Dundee, UK)
- Matthew Purver (Queen Mary University of London, UK)
- Georgi Stojanov (American University of Paris, France)
- Hannu Toivonen (University of Helsinki, Finland)
- Tony Veale (University College Dublin, Ireland)
- Dan Ventura (Brigham Young University, USA)
- Geraint Wiggins (Queen Mary, University of London, UK)

The Unnoticed Creativity Revolutions: Bringing Problem-Solving Back into Computational Creativity

Tarek R. Besold [1]

Abstract. Computational creativity is steadily gaining popularity and has become a recognised field of scientific activity. Still, while work in art-performing and artefact-producing computational creativity has greatly advanced, research into the computerisation of other forms of creativity, such as general creative capabilities for computational cognitive agents or automated creative problem-solvers addressing real-world scenarios in a domain-independent way, is lacking far behind and is receiving only limited attention from the community. In fact, over the last years other disciplines have been reporting developments which—when looked at from the perspective of computational creativity—could turn out to be crucial stepping stones for advancing towards closing this gap between "artistic creativity" and "problem-solving creativity" but which went mostly unnoticed by the computational creativity community.

In this paper I will have a closer look at the differences between artistic (computational) creativity and problem-solving (computational) creativity, followed by a review of two developments from cognitive modelling and machine learning which serve as cases in point for breakthroughs with potentially high relevance for problem-solving computational creativity research. I will then use these examples to motivate the claim that for both individual researchers as well as the computational creativity community as such the time has come to also focus on problem-solving creativity.

1 Computational Creativity in Early 2016: The Disparate Success of Different Lines of Research

Computational creativity originally started as part of research in Artificial Intelligence (AI) but as of today has become a multidisciplinary endeavour spanning from computer science to topics in cognitive psychology, philosophy, and the arts (an overview of the history of the field by one of its founders—tracing the vision of creatively active computing machines back at least to Ada Lovelace—can, for instance, be be found in [8]). While this multidisciplinary nature makes it hard to precisely define which scientific questions and research endeavours count into the subject area of computational creativity as scientific discipline, one of the fairly widely accepted working definitions has been suggested in [11] as:

> "The philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative."

Looking at the actual practice in the field and abstracting from it, the goals of most computational creativity research projects can be unified under the overarching idea of the modelling, simulation, or replication of creativity using computational means in order to achieve one of several ends: Either to construct a program or computer with human-level creative capacities, or to better understand human creativity and to formulate a (in the Marrian sense [19]) computational- or algorithmic-level perspective on creative behaviour in humans, or to design co-creative application programs that can contribute to or enhance the creativity of their (human) users without necessarily being creative themselves. In way of a quite general summary computational creativity therefore concerns itself with theoretical and practical issues in the study of creativity. Theoretical work on the nature and proper definition of creativity (with a strong emphasis towards aspects which are also relevant for or to be found in computationally creative systems) is conducted side by side with practical work on the implementation of computational modules (within more complex and/or general architectures) or entire stand-alone systems that exhibit creativity, with both strands of work being in continuous exchange and cross-informing and -influencing each other.

The popularity of computational creativity as an academic field in its own right (i.e., having a community with its members subsuming their respective research efforts under the umbrella term computational creativity, and with academic meetings, publication outlets, and funding lines specifically dedicated to the corresponding scientific agendas) has steadily been growing since the First Joint Workshop on Computational Creativity in 2004: Last year, the International Conference on Computational Creativity (the academically matured successor series of events of the Joint Workshops on Computational Creativity) witnessed its sixth edition ICCC-2015, since 2013 four still ongoing EU-funded FP7 FET-Open research projects are explicitly dedicated to theoretical and practical questions in computational creativity, and the FP7 three-year coordination action "Promoting the Scientific Exploration of Computational Creativity (PROSECCO)"[2] has been actively promoting and raising awareness for computational creativity with other scientific communities and with the general public. Scientifically, papers reporting on work in computational creativity are by now part of the standard repertoire of every major AI conference and frequently appear in high-ranking journals, and the International Joint Conference on AI (IJCAI) as flagship conference of the entire discipline even featured a dedicated track on "AI and the Arts" as part of its latest edition in Buenos Aires in summer 2015. Also, several journals have published special issues on different aspects of computational creativity (see, for example, [10, 4, 2]), with [3] the first book reporting exclusively on research conducted in the field of computational creativity has been published, and several other collections have dedicated a significant share of attention to the topic [21, 35].

With respect to the achieved results, several successes of computa-

[1] The KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano, email: TarekRichard.Besold@unibz.it

[2] `http://prosecco-network.eu`

tional creativity have been reported (and also publicly acknowledged by the wider public) from the arts: artificial systems have been accepted as creators of artistic artefacts (and, thus, potentially as creative) by the corresponding audience. In order to just name two examples, musical compositions by the University of Malaga's Iamus system [13] have been performed and recorded by the London Symphony Orchestra and other world-famous artists and ensembles, and paintings produced by Cohen's AARON [20] have been sold in the standard art market since the 1970s. Moreover the scope is not limited to visual or musical forms of artistic expression but also covers, for instance, the culinary arts. IBM's Watson system has successfully been adapted to produce novel and interesting cooking recipes [28], the outcomes of which have been selected and prepared by internationally leading chefs and featured at different festivals and receptions. And also some of the creations of [36]'s PIERRE system have found positive reception when served at the "*You Can't Know My Mind*" creativity festival.

Still, the situation looks far less rosy on the more cognitive systems and problem-solving oriented side (i.e., the line of work trying to imbue computational cognitive agents with general creative capacities or to develop systems capable of addressing practical creativity challenges in real-world scenarios and/or mostly independent of the concrete domain). Here, progress has been slower and many questions concerning the cognitive nature and computational modelling of creativity (for example in inventive problem-solving, automated theory development, or concept invention) remain unanswered or—more often than one would prefer—even unasked.

This delay in development is strongly related to the question for a general definition of creativity as cognitive capacity as one of the fundamental challenges in creativity research and computational creativity, and also to the community's approach to this core problem. While humans usually fairly immediately and unmediatedly recognise (or at least are willing to indicate an opinion about) the presence or absence of creativity in different forms of artistic performance, in a problem solution, or in the execution of a task, attempts at giving an explicit characterisation of creativity or a set of (reasonably general) sufficient criteria for deciding when an artefact, behaviour, or idea should be judged as creative hitherto have not succeeded. And even when asking for necessary criteria the situation presents itself as fairly bleak with most suggestions falling significantly short of providing a conclusive—or at least directly practically usable—answer. Take as an example the following often quoted characterisation of creativity from [25]:

> "Over the course of the last decade, however, we seem to have reached a general agreement that creativity involves the production of novel, useful products."

This is problematic in at least two ways. On the one hand the account is already restricted to a subset of the huge number of different forms of creative activity (namely those manifestations involving the production of some type of artefact), on the other hand both novelty and usefulness as necessary requirements are at the same time highly underspecified and in their possible interpretations strongly dependent on specific contexts and domains. Similar criticism can be put forward for most (if not all) other attempts at characterising creativity, which often are similar in approach to Mumford's quoted suggestion (see, for instance, the description of creative people as agents creating "something original and worthwhile" in [31]). Unfortunately, it additionally seems as if a remedy to this unsatisfactory—and particularly for research in computational creativity methodologically quite challenging—situation cannot be expected any time soon; the computational creativity community at present has mostly abandoned the search for a generally acceptable and practically usable definition of creativity, having grown weary of what—as so often in the process of finding a definition for a phenomenon as multi-facetted as creativity—seemed to be almost undecidable discussions proceeding in at best infinitesimally small steps.

Coming back to the two distinct general strands of research in computational creativity mentioned before, the lack of proper definition or characterisation impacts on them to quite different degrees. In the case of artefact-producing or art-performing systems in most current systems emphasis and attention are almost exclusively focused on the output and potentially accompanying artistic narratives (generated by the computational system in order to provide the audience with some form of a "creation story" mimicking a human artist's explanation of the genesis and content) rather than on the actual computational and algorithmic process details driving the production or performance. This has the advantage that the output can be evaluated and subsequently used for guiding the system's further development.[3] Alas this doesn't hold true for the more process-focused problem-solving side of the research spectrum. The latter is significantly impaired by the lack of definitions or characterisations of creativity tangible enough to serve as basis for process models or mechanistically-informative theories which then could provide a foothold for a computational implementation of general creative capabilities in a cognitive agent or in a domain-independent problem-solver. This impasse and the resulting lack of progress—together with certain sociological and science-philosophical considerations in relation to the AI community as traditional home of research into problem-solving, cognitive agents, and the like—resulted in many researchers in computational creativity dedicating their efforts rather to other projects and topics than pursuing what seemed to be an arduous challenge with unclear chances of success, from the computational creativity perspective leaving the scenery unchanged and mostly unattended over the last years.

Still, precisely due to the multidisciplinary nature of computational creativity many interests and challenges are shared with other fields of research, sometimes openly and explicitly but more often disguised behind different terminologies and looked at from different perspectives. So while computational creativity has treated problem-solving and general creativity in cognitive agents as an orphan of kinds other disciplines have witnessed remarkable breakthroughs in questions relating to these topics, showing that progress is possible also within this strand of work.

In what follows I will first have another look at the relation be-

[3] It has been argued that the degree of creativity of a computational system cannot only be judged by its outputs, but that the assessment has to take into account several different factors which might not necessarily be reflected in the output alone. For instance in [9] the suggested aspects are skill, appreciation, and imagination each evaluated from the perspective of the programmer, the computational system, and the consumer, resulting in a model with a 3×3-dimensional factor space. Still, while this and similar complex models might be required for a comprehensive assessment of a system's creative potential from the academic perspective of computational creativity research, I strongly believe that in practice the evaluation of a computationally creative system by non-experts is mostly based on the system's output (especially when additionally providing an accompanying artistic narrative dissuading the audience from enquiring about the process). And also from a computational creativity practitioner's perspective assessment measures taking into account properties of the process pose significant challenges and open questions—such as "Does using random processes for creating non-deterministic system behaviour really reflect imagination?" or "Does the fitness function of an evolutionary art system really reflect appreciation of the system's own work?"—making output-based approaches to evaluation appear simpler and more straightforward.

tween creativity in artefact-production and artistic performances on the one hand and problem-solving and general practical cognitive agency on the other hand, pointing out several key features setting the two categories apart. Sect. 3 then showcases two example breakthroughs (taken respectively from cognitive science and machine learning) other disciplines have made with respect to potentially crucial aspects of the latter type of creativity. Considerations and recommendations on how to bring research on problem-solving related creativity back into computational creativity are given in Sect. 4. Sect. 5 concludes the paper with a short summary of the main points made.

2 Further Distinctions Between Artistic and Problem-Solving (Computational) Creativity

Several categorisations for the many different facets of creativity have been proposed. Using the degree of creativity on an agent-centric biographic or cultural scale as measure, some researchers distinguish between day-to-day creative acts as "little-c" creative and stroke-of-genius creative moments as "Big-C" creative [18], while others draw a conceptually quite similar line between "p-creative" ideas, as ideas which are new for an individual (i.e., a person coming to a new realisation), and "h-creative" ideas which are historically new (i.e., genuinely novel and unprecedented inventions) [6]. Additionally—in parts complementing, in parts completing the bimodal view of creativity—with "mini-c" creativity inherent in learning and personal discovery processes (as, for example, all the time happening during children's development and education) and "Pro-c" creativity as the developmental and effortful progression beyond little-c (representing professional-level expertise in any creative area) at least two more forms of creativity have been identified [17]. Alternatively, taking a more process type-oriented perspective, a distinction between combinatorial, exploratory, and transformational creativity has been suggested [7, 5] (a formalization of which has been attempted in [38]). Combinatorial creativity is taken as creativity based on combining pre-existing ideas or objects, exploratory creativity corresponds to searching for novel concepts within an area of a conceptual space—where concepts are taken as locations in the space—which adheres to certain rules, and transformational creativity is thought to alter these rules and, thereby, create a new area of conceptual space to be searched in.

While these categorisations aim at distinguishing between different forms of creativity on the object level of creativity as manifested in the world, within computational creativity as discipline aiming at making creativity computationally accessible an additional meta-layer of characterisation and analysis becomes relevant regarding the nature of creativity as computational phenomenon. The question on the meta-level is captured—following the spirit of Searle's classical distinction of the two conceptual approaches to Artificial Intelligence [29]—in the distinction between "weak" and "strong" computational creativity described in [1]:

> "'[W]eak computational creativity', which does not go beyond exploring the simulation of human creativity; emphasising that genuine autonomy and genuine understanding are not the main issues in conceptualising weak computationally creative systems. Conversely in 'strong computational creativity', the expectation is that the machine should be autonomous, creative, have 'genuine understanding' and other cognitive states."

I want to introduce a third classification cutting across both the object level and the meta-level of computational creativity. In the diagnosis of the state of affairs within the two different strands of work in computational creativity in Sect. 1 a distinction was made between art-performing and artefact-producing manifestations of creativity on the one hand, and the process-focused problem-solving form of creativity on the other hand. As already hinted at in the discussion of possible reasons for the quite disparate level of development and success in respectively computerising them, this distinction seems to cut deeper than merely being situated at the level of published papers or press releases. Instead it rather appears to have its roots in qualitative differences between both forms of creativity and between the required computational paradigms and methodologies for each. Besides manifesting in the ways already described in the opening section, the distinction becomes further obvious when considering its relation to the just introduced categorisations of creativity and computational creativity.

Still, before relating the newly proposed division to the pre-existing classifications for the sake of simplicity I want to introduce two class labels for the new categories. In what follows I will refer to artefact-producing and art-performing forms of creativity as "artistic creativity" as opposed to "problem-solving creativity" as second family of manifestations of creativity. For the latter category also "practical creativity" or "pragmatic creativity" could have been used as name as the coverage shall not be restricted to classical (fairly artificial) problem-solving scenarios as known from the AI literature, such as the Tower of Hanoi problem [26], or (purposefully contrived) explicit insight problems, such as the nine dots puzzle [14]. Instead, a wide range of activities and processes involving creativity shall be summarised under this label including, but not limited to, inventive problem-solving in applied settings, automated theory development about arbitrary domains, or concept invention in real-world scenarios. Still, "problem-solving creativity" seems to be the best covering term for this diverse set of expressions of creativity as all corresponding occurrences can be phrased in the corresponding conceptual framework with problem description, starting situation, and the outcome of the creative process as potential solution.[4]

The distinction between artistic and problem-solving creativity is different from the previously cited categorisations of creativity on the object level since the property used for classification is not focusing on either the degree of creativity or the type of creative process but arises from the type of manifestation of creativity, distinguishing—very roughly speaking—between arts-related forms of creativity with strong emphasis on production aspects and creativity with strong emphasis on action in or conceptualisation of a creative agents environment and lived-in world. Therefore the differentiation between artistic and problem-solving creativity is at first instance ignorant especially to the biographic or cultural dimension and degree of creativity, making it conceptually independent from notions of h-creativity, p-creativity, and the like. A practice-mediated correlation is shared with the process type-centered approach to classification, though. While both ways of characterising do not have a direct conceptual connection in their classification criteria, when having a look at how most prominent systems in artistic creativity operate their approach to creativity can be considered combinatorial or exploratory with (attempts at) transformational creativity—while existent—being the occasional exception.[5] For the case of problem-solving creativity it

[4] Using the term "practical creativity" instead runs the risk of excluding manifestations of creativity which seem not to directly produce a practical outcome as, for instance, many cases of scientific theory formation. "Pragmatic creativity" on the other hand seems to predesignate exclusively the real-worldly impact or the success of the resulting action of the creative process as measure of assessment and benchmark for creativity resulting in unsolicited limitations.

[5] For the case of artistic computationally-creative systems in the music do-

can be expected that–while also combinatorial and exploratory aspects will be indispensable—transformational creativity will play a much more important role. Studies in cognitive psychology and cognitive science suggest that cross-domain reasoning processes, such as analogy-making [16] and concept blending [15], as well as meta-reasoning processes [12] play an important role in many different forms of problem-solving, learning (i.e., concept acquisition or invention), and other creative activities very frequently performed by cognitive agents. These processes are genuinely transformational in nature, importing qualitatively different rules and content into the conceptual space or making the underlying rules of the space itself (as opposed to rules defined within the space) accessible and modifiable, thereby altering the entire space.

When looking at the meta-level distinction between weak and strong computational creativity, the classification in artistic or problem-solving creativity *a priori* again does not have a necessary direct conceptual relation or correspondence. Still, again a practice-mediated correlation can be expected. While current computational systems performing artistic creativity are of the weakly creative type according to the above-cited definition from [1], problem-solving creativity seems to be moving closer to the realm of strong computational creativity. Transformational creativity inevitably presupposes some form of access to and handle on the governing characteristics of the conceptual space—for instance meta-reasoning in most accounts presupposes a certain (self-)recognition of defining principles of the reasoner and of its environment as basis for the required meta-abstraction—and most theories and models of cross-domain reasoning processes are inspired by cognitive capacities and theories about cognitive states and their relations. This does not mean that I deem cognitive states or human-like understanding a necessary criterion for building cognitive agents with fairly general problem-solving and creative abilities as many creative feats which humans perform drawing on their cognitive states and human cognitive qualities might well be achievable in different ways using computational means. Still, I am certain that growingly complex and powerful computational systems for problem-solving creativity will eventually approach and push our expected boundary between weak and strong creativity, in all likelihood forcing us to reconsider many of our initial intuitive ideas about the limitations of weak computational creativity. On the other hand I explicitly do not want to rule out the possibility that for a general computational solution to problem-solving creativity strong computational creativity would be required, possibly making full problem-solving creativity another example instance of the class of AI-complete problems [22, 30].

In summary, the distinction between artistic and problem-solving creativity does not seem to be equivalent to any of the discussed earlier categorisations, while still being based on a genuine qualitative difference between both forms of creativity and between computational approaches to them. Also, this qualitative difference is not reducible to the mere difference in type of manifestation or domain as main criterion for classification, but has significantly farther reaching ramifications. Now, closing the circle and looking at the disparity in level of development and success of research work between both families—with work in artistic computational creativity flourishing while there are only few efforts within the computational creativity community towards tackling problem-solving creativity—the question arises of whether the quest for methods and approaches relevant for computational accounts of problem-solving creativity has really been abandoned? Or are the corresponding challenges simply being addressed by other communities outside computational creativity in a way mostly unnoticed or unrecognised by the field?

main the border between exploratory and transformational creativity may seem less clear than, for instance, in computational accounts of artistic painting. Still, as most musical systems are based on some form of rule- or constraint-driven approach using, for example, generative grammars or constraint satisfaction methods, removing some of the respective rules or constraints during computation in my opinion does not yet constitute a change from exploratory to transformational creativity but rather widens the area of search from a restricted area within one conceptual space to a bigger sub-area.

3 Creative Computing Outside Computational Creativity: Where Problem-Solving Went

As stated towards the end of Sect. 1 many questions in computational creativity are shared with other fields of research. Clearly, there often will be differences in terminology and variations in the conceptualisations of problems due to different overarching themes, different histories, and different standard methods and approaches, but once these particularities of the concrete instantiation of the corresponding research question have been stripped away the underlying structure and basic interest coincide.

Since most research in computational creativity currently is dedicated to artistic creativity and problem-solving creativity plays only very minor walk-ons on the academic stage, for fans of the latter it is worth to have a look at some related fields of scientific investigation and see whether there has been any progress which could be relevant for problem-solving creativity once carried over to computational creativity. In order to substantiate this claim in this section I accordingly have a cursory look at cognitive science as field which, among others, also studies creativity and at inductive logic programming as subfield of machine learning within AI which also deals with concept acquisition and invention. While doing so I try to identify current developments in these disciplines which could be highly relevant for computational approaches to problem-solving creativity and which might contribute to finally advancing again also in this line of computational creativity. Without any doubt the following examples are far from representing an exhaustive list of relevant developments within these disciplines, as not even to mention other fields where relevant research is surely happening, too. They are rather arbitrarily chosen cases in point which I hope will serve as motivation and incentive for broadening and intensifying the search and for starting to carry over promising notions into computational creativity in order to test their actual potential for applications in problem-solving creativity.

3.1 Bayesian Models of Theory Learning in Cognitive Science

A field directly involved in the study of creativity—although focused on the human version thereof—is cognitive science. Cognitive science also unites different disciplines under its interdisciplinary roof, among others having close ties to AI and computer science. When searching for the currently dominant topics in cognitive science one of the clear leading paradigms is cognitive modelling using theory-based Bayesian models [33]. Having started out as an attempt to understand human inductive learning and reasoning in computational terms by building fairly general, quantitatively predictive models approximating optimal inference in natural environments, Bayesian models of theory learning have now been developed for a huge variety of tasks and application scenarios. The basic idea underlying the modelling paradigm is to cast induction as a form of Bayesian statistical inference over structured probabilistic models of the world, which can be seen as probabilistic versions of intuitive theories providing the required domain knowledge for inductive generalisation from sparse data.

Among the many different Bayesian models on the market are also approaches to theory acquisition from observed data. For example in [34] a model covering the development of children?s intuitive theories about simple physical properties such as magnetism is given. There it is assumed that a domain of cognition is given which provides the learner with some observed data. The task now is to build a theory of the domain given these observations, i.e., to find a set of abstract concepts and explanatory laws that together generate a hypothesis space and prior probability distribution over candidate models for systems in that domain. When applied to the magnetism example in [34] the concrete situation can be described as follows:

> "For example, consider a child learning about the domain of magnetism. She might begin by playing with a few pieces of metal and notice that some of the objects interact, exerting strange pulling or pushing forces on each other. She could describe the data directly, as "Object i interacts with object f", "Object i interacts with object j", and so on. Or she could form a simple theory, in terms of abstract concepts such as magnet, magnetic object and non-magnetic object, and laws such as "Magnets interact with other magnets", "Magnets interact with magnetic objects", and "Interactions are symmetric" (but no other interactions take place). (...) A model of a system specifies the minimal facts needed to apply the abstract theory to the system, in this case which objects are magnetic, which are magnets, and which are non-magnetic. From these core facts the laws of the theory determine all other true facts—in our example, this means all the pairwise interactions between the objects: e.g., objects i and j, being magnets, should interact, but i and e should not, because e is non-magnetic. Finally, the true facts generate the actual data observed by the learner via a noisy observation process."

The Bayesian model in [34] (also see the conceptual sketch in Fig. 1) constitutes not only a model in the general sense of children's cognitive development but also provides a model specifically of a mini-c creative process in the sense of [17], describing a cognitive agent's process of "making sense" of its environment and in the process discovering a simple theory of magnetism (which at the same time can be seen as an occurrence of p-creativity). This capacity of generating theories from observations has to be considered crucial for problem-solving creativity in at least two ways: As creative manifestation by itself, for example for automated theory development as application of computational creativity in science, and as stepping stone towards the ability of a computational system to use observations for identifying the defining and governing rules of any problem domain or the system's environment—which in turn is a necessary prerequisite to targeted and guided transformationally-creative processes.

3.2 Meta-Interpretive Learning and Predicate Invention in Inductive Logic Programming

The basic question in Inductive Logic Programming (ILP) [27] is quite similar to the setting described in the previous subsection on Bayesian theory learning. In the most general ILP scenario, given some background knowledge and a set of examples represented as a logical database of facts, the computational system is tasked to derive a hypothesised logic program which entails all the positive and none of the negative examples from the database. In doing so also a

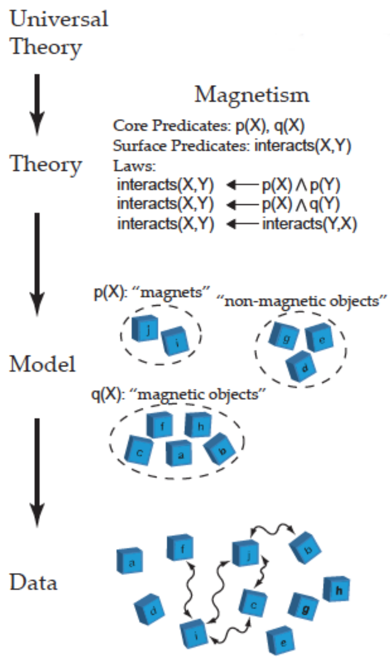
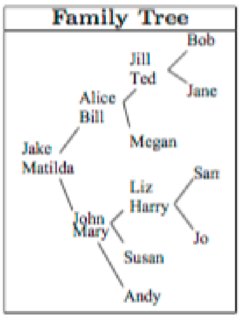


Figure 1: The Magnetism example for Bayesian theory learning from [34]: Assuming a hypothesis space of possible theories constrained by some Universal Theory, the learner has to find a set of abstract concepts and explanatory laws (i.e., a theory) that together generate a hypothesis space and prior probability distribution over candidate models, with each model generating a probability distribution over possible observations.

form of theory learning is taking place with the theory being represented in form of the logic program which is taken as "explaining" the observed facts.

Lately, ILP has started to regain attention within AI research since a long-standing problem has finally been overcome: efficient methods for predicate invention and for the learning of recursion have been presented. Consider the following example (taken from [24]): Given the Family Tree in Fig. 2 and the task for the system to learn the corresponding logic program. Now, since examples of the ancestor relation are provided while the background knowledge only lists facts concerning father and mother, the system must perform two tasks in parallel, namely learn the recursive definition of ancestor and at the same time invent parent in order to actually make this learning process possible in the first place. Only then can the target theory be obtained. In [23], Meta-Interpretive Learning (MIL) has been introduced. MIL is a technique which supports efficient predicate invention and learning of recursive logic programs built as a set of metalogical substitutions by a meta-interpreter which acts as the central part of an ILP learning engine. The user provides the meta-interpreter with higher-order expressions defining the permitted forms of clauses admissible in the hypothesised programs as meta-rules. The meta-interpreter then attempts to prove the respective examples and, for any successful proof, saves the substitutions for existentially quantified variables found in the associated meta-rules. When these substitutions are applied to the meta-rules they result in a first-order definite program which is an inductive generalisation of the examples.



Target Theory
$father(ted, bob) \leftarrow$
$father(ted, jane) \leftarrow$
$parent(X, Y) \leftarrow mother(X, Y)$
$parent(X, Y) \leftarrow father(X, Y)$
$ancestor(X, Y) \leftarrow parent(X, Y)$
$ancestor(X, Y) \leftarrow parent(X, Z), ancestor(Z, Y)$

First-order	Metalogical substitutions
Examples $ancestor(jake, bob) \leftarrow$ $ancestor(alice, jane) \leftarrow$	N/A
Background Knowledge $father(jake, alice) \leftarrow$ $mother(alice, ted) \leftarrow$	N/A
Instantiated Hypothesis $father(ted, bob) \leftarrow$ $father(ted, jane) \leftarrow$ $p1(X, Y) \leftarrow father(X, Y)$ $p1(X, Y) \leftarrow mother(X, Y)$ $ancestor(X, Y) \leftarrow p1(X, Y)$ $ancestor(X, Y) \leftarrow p1(X, Z), ancestor(Z, Y)$	 $metasub(instance, [father, ted, bob])$ $metasub(instance, [father, ted, jane])$ $metasub(base, [p1, father])$ $metasub(base, [p1, mother])$ $metasub(base, [ancestor, p1])$ $metasub(tailrec, [ancestor, p1, ancestor])$

Figure 2: The Family Tree example for predicate invention and meta-interpretive learning in ILP from [24]: Upper-left the actual family tree, upper-right a corresponding target theory (i.e., logic program) including two examples in the first two lines. The left column in the lower half features (from top to bottom) two observations from the domain, two entries from the background knowledge database, and a hypothesised logic program constructed from instantiated meta-rules listed in the right column.

In this way, given an appropriate meta-interpreter, the two examples $father(ted, bob)$ and $father(ted, jane)$ in the upper part of Fig. 2 can be obtained.[6] Also, Fig. 2 provides an example for successful predicate invention: The predicate $p1$ encodes the real-world family relationship parent required for learning the definition of ancestor.

Again a mini-c and p-creative process has taken place and the system has performed an act of concept invention, creating a predicate corresponding to the concept of parent and thereby genuinely expanding its space of learnable relations (as without the parent predicate the ancestor relationship which is recursively defined over parent structures could not be obtained). As had been the case with Bayesian theory learning, also the capacity to create previously unseen and unrepresented concepts (i.e., concept invention) from observations and knowledge about a domain can certainly be assumed to be crucial for implementing models of problem-solving creativity, both in particular application contexts as well as on the general level of accessing and implementing transformational creativity.

4 Ways and Reasons for Bringing Problem-Solving Back into Computational Creativity

The examples in the previous section highlight research work outside of computational creativity which carries the promise to be highly relevant in systems aiming at accessing transformational creativity and at implementing creative aspects of general cognitive agency in a real-world setting. And although the described methods have been developed in completely different contexts and in all likelihood without any thought relating to computational creativity, their potential relevance for problem-solving creativity becomes strikingly apparent once the connection has been established. Bayesian theory learning and predicate invention/MIL are only two cases among what I strongly believe to be a surprisingly high number of revolutionary developments in other areas within the domain of computer science and related areas. But while these breakthroughs—once identified and properly re-contextualised—also have the potential to turn out as revolutions from the computational creativity point of view they hitherto simply have gone unnoticed by the big majority of researchers currently active in computational creativity.

I am convinced that a thorough review of recent advances in disciplines such as machine learning, commonsense reasoning, cognitive systems, cognitive robotics, and computational cognitive modelling will quite quickly uncover many results which can directly be converted and re-applied in problem-solving creative applications. This would also have the advantage that an approach based on transfer and adaptation of already tested techniques and implementations probably shortens the required time and reduces the risk of failure for getting from the idea to an operational proof of concept system as compared to a development starting from scratch. While this might—in my opinion completely unjustifiedly—seem scientifically less prestigious to some, shorter development cycles with only limited risk promise initial successes to happen in a timely manner. These successes in turn will be proof that advances are also possible on the side of problem-solving creativity motivating researchers to return to

[6] See [24] for details concerning overall MIL setting, as well as regarding the specific meta-interpreter and the generation of the hypotheses in the example.

this strand of scientific investigation and to thereby put work on the corresponding topics and questions back on the map.

To me this would also be desirable from the perspective of computational creativity as scientific community. The many-faceted phenomenon of creativity clearly is part of human cognition and of the way humans as cognitive agents interact with their environment. Still, this spectrum of interaction spans far beyond the realm of artistic activity and performance and basically covers all domains of our daily lifeworld. Work in artistic computational creativity is important and the achieved recognition is well-deserved, but this should not limit the focus of our community—which carries "creativity" as general, unqualified term in its self-given name—to certain forms of creativity whilst ignoring others. AI for a few years now is witnessing a period of reinvigoration and rejuvenation, receiving a high level of public attention. This is in parts a self-reinforcing process, since each reported success entices more researchers to invest themselves in the field. And indeed, while many of the long-standing problems at the core of AI still remain unanswered, solutions to some others have been presented and at first sight promising approaches to others have been suggested. For example cognitive systems and cognitive robotics (but also other disciplines) are advancing at a pace unseen for the previous two decades, and the addressed questions while still quite basic are coming closer to also addressing (parts of) phenomena such as creativity. I am confident that these communities could greatly profit from interacting with the field of computational creativity and its practitioners, drawing on the latter's knowledge and experience when starting to address creativity-related questions in their respective research endeavours. And as argued before, the exchange would be reciprocal since also computational creativity—and especially the problem-solving creative line of investigation—certainly could greatly profit from developments in other disciplines of AI and related areas.

5 Conclusion

In the previous sections I have first taken stock of the field of computational creativity as it stands in early 2016, finding that work in computerising artistic creativity has shown great success and receives significant attention by researchers and the public while work on problem-solving creativity has currently been relegated to an existence at the sideline of investigation and community activity. This prompted me to have a closer look at possible reasons for this state of affairs and, subsequently, at the differences between both forms of creativity and the corresponding research efforts. Problem-solving creativity turned out to have closer ties to transformative creativity and to be computationally closer to (in the sense of [1]) strong computational creativity than is generally the case for artistic (computational) creativity. I then gave two examples for developments from other fields with potentially high relevance for problem-solving computational creativity, using these examples as empirical motivation for the claim that notwithstanding the just stated characteristics advances in questions relating to problem-solving creativity are possible and actually are happening. Based on this observation I then argued in favour of initiating a process bringing problem-solving creativity back into the center of computational creativity research based on both, considerations regarding the scientific questions and content as well as reflections concerning computational creativity as a community and its relation to AI and related fields.

One thing left to do is to point out that current computational creativity research is not entirely ignorant of problem-solving creativity and that examples of research into aspects of problem-solving creative systems can be found. One instance for such a project is the development of the "Information Dynamics of Thinking" (IDyOT) computational cognitive architecture reported in [37]. IDyOT attempts to provide a simple and functional cognitive model and proof of the claim that information-theoretically regulated probabilistic prediction—which is hypothesised to serve as a general mechanism for managing information and events in the world—is the mechanism underlying non-conscious creativity. And also work in cognitive architectures like the project described in [32], trying to computationally model and simulate the psychological processes underlying human creative problem solving using the CLARION architecture, clearly has to be counted as important contribution to the study of problem-solving (computational) creativity. Still, the number of examples for research endeavours of these and similar types is very limited and clearly too small when compared to the possible impact of successfully solving any significant part of problem-solving creativity.

Based on what I previously discussed, in closing I want to invite researchers from computational creativity and also from AI and related fields in general to consider work on problem-solving computational creativity. I am convinced that many opportunities for development and scientific progress are currently left unused in this area, waiting to be recognised and converted into a new generation of computationally creative systems.

Acknowledgements

The author acknowledges the financial support of the Future and Emerging Technologies within the 7th Framework Programme for Research of the European Commission, under FET-Open grant 611553 (COINVENT).

REFERENCES

[1] M. M. al Rifaie and M. Bishop, 'Weak and strong computational creativity', in *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, 37–49, Atlantis Press, (2015).
[2] T. R. Besold, K.-U. Kühnberger, and T. Veale, *Special Issue: Computational Creativity, Concept Invention, and General Intelligence*, volume 6(1) of *Journal of Artificial General Intelligence (JAG)*, 2015.
[3] *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, Atlantis Press, 2015.
[4] M. Bishop and Y. J. Erden, *Special Issue: Computational Creativity, Intelligence and Autonomy*, volume 4(3) of *Cognitive Computation*, 2012.
[5] M. Boden, 'Computer models of creativity', *The Psychologist*, **13**, 72–76, (2000).
[6] M. A. Boden, *The Creative Mind: Myths and Mechanisms*, Basic Books, 1990.
[7] M. A. Boden, 'Creativity and artificial intelligence', *Artificial Intelligence Journal*, **103**, 347–356, (1998).
[8] M. A. Boden, 'Foreword: How computational creativity began', in *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, v–xiii, Atlantis Press, (2015).
[9] S. Colton, 'Creativity versus the perception of creativity in computational systems', in *AAAI Spring Symposium: Creative Intelligent Systems*, volume SS-08-03 of *AAAI Technical Reports*, 14–20, (2008).
[10] S. Colton, R. Lopez de Mantaras, and O. Stock, *Computational Creativity: Coming of Age*, volume 30(3) of *AI Magazine*, 2009.
[11] S. Colton and G. Wiggins, 'Computational creativity: The final frontier?', in *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI)*, pp. 21–26. IOS Press, (2012).
[12] *Metareasoning: Thinking about Thinking*, eds., M. T. Cox and A. Raja, MIT Press, 2011.

[13] G. Diaz-Jerez, 'Composing with melomics: Delving into the computational world for musical inspiration', *Leonardo Music Journal*, **21**, 13–14, (2011).
[14] G. T. Dow and R. E. Mayer, 'Teaching students to solve insight problems. Evidence for domain specificity in training', *Creativity Research Journal*, **16**(4), 389–402, (2004).
[15] G. Fauconnier and M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*, Basic Books, 2002.
[16] *The analogical mind*, eds., D. Gentner, K. J. Holyoak, and B. N. Kokinov, MIT Press, 2001.
[17] J. Kaufman and A. Beghetto, 'Beyond Big and Little: the Four C Model of Creativity', *Review of General Psychology*, **13**(1), 1–12, (2009).
[18] *The international handbook of creativity*, eds., J. Kaufman and R. Sternberg, Cambridge University Press, 2006.
[19] J. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco, CA, 1982.
[20] P. McCorduck, *Aaron's code: meta-art, artificial intelligence, and the work of Harold Cohen*, Macmillan, 1991.
[21] *Computers and Creativity*, eds., J. McCormack and M. d'Inverno, Springer, 2012.
[22] E. T. Mueller, 'Daydreaming and computation: A computer model of everyday creativity, learning, and emotions in the human stream of thought', Technical Report CSD-870017, University of California, Los Angeles, (March 1987).
[23] S. H. Muggleton, D. Lin, N. Pahlavi, and A. Tamaddoni-Nezhad, 'Meta-interpretive learning: application to grammatical inference', *Machine Learning*, **94**, 25–49, (2014).
[24] S. H. Muggleton, D. Lin, and A. Tamaddoni-Nezhad, 'Meta-interpretive learning of higher-order dyadic datalog: predicate invention revisited', *Machine Learning*, **100**, 49–73, (2015).
[25] M. D. Mumford, 'Where have we been, where are we going? taking stock in creativity research', *Creativity Research Journal*, **15**, 107–120, (2003).
[26] A. Newell and H. A. Simon, *Human problem solving*, Prentice-Hall, Englewood Cliffs, NJ, 1972.
[27] S.-H. Nienhuys-Cheng and R. de Wolf, *Foundations of Inductive Logic Programming*, volume 1228 of *Lecture Notes in Artificial Intelligence*, Springer, 1997.
[28] F. Pinel, L. R. Varshney, and D. Bhattacharjya, 'A culinary computational creativity system', in *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, 327–346, Atlantis Press, (2015).
[29] J. Searle, 'Minds, brains, and programs', *Behavioural and Brain Sciences*, **3**, 417–457, (1980).
[30] S. C. Shapiro, 'Artificial intelligence', in *Encyclopedia of Artificial Intelligence (2 ed.)*, ed., S. C. Shapiro, 54–57, John Wiley, (1992).
[31] R. J. Sternberg and K. Sternberg, 'Problem solving and creativity', in *Cognitive Psychology (6 ed.)*, Cengage Learning, (2011).
[32] R. Sun and S. Helie, 'Accounting for creativity within a psychologically realistic cognitive architecture', in *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, 151–165, Atlantis Press, (2015).
[33] J. B. Tenenbaum, T. L. Griffiths, and C. Kemp, 'Theory-based bayesian models of inductive learning and reasoning', *Trends in Cognitive Science*, **10**, 309–318, (2006).
[34] T. D. Ullman, N. D. Goodman, and Joshua B. Tenenbaum, 'Theory Acquisition as Stochastic Search', in *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, (2010).
[35] *Creativity and the Agile Mind: A Multi-Disciplinary Study of a Multi-Faceted Phenomenon*, eds., T. Veale, K. Feyaerts, and Ch. Forceville, volume 21 of *Applications of Cognitive Linguistics*, De Gruyter Mouton, 2013.
[36] D. Ventura, 'The computational creativity complex', in *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, 65–92, Atlantis Press, (2015).
[37] G. A. Wiggins and J. Forth, 'IDyOT: A Computational Theory of Creativity as Everyday Reasoning from Learned Information', in *Computational Creativity Research: Towards Creative Machines*, eds., T. R. Besold, M. Schorlemmer, and A. Smaill, volume 7 of *Atlantis Thinking Machines*, 127–148, Atlantis Press, (2015).
[38] Geraint A. Wiggins, 'A preliminary framework for description, analysis and comparison of creative systems', *Knowledge-Based Systems*, **19**(7), 449–458, (November 2006).

Swarmic Autopoiesis: Decoding de Kooning

Mohammad Majid al-Rifaie[1] and **William Latham** and **Frédéric Fol Leymarie** and **Mark Bishop**

Department of Computing, Goldsmiths
University of London, London SE14 6NW
United Kingdom,
Emails: {m.majid, w.latham, ffl, m.bishop} @gold.ac.uk

Abstract. This paper introduces a novel approach deploying the mechanism of 'attention' by adapting a swarm intelligence algorithm (Stochastic Diffusion Search) to selectively attend to areas of a digital canvas (with line drawings) that has more details (e.g. lines or points). Once the attention of the swarm is drawn to a certain line within the canvas, the capability of another swarm intelligence algorithm (Particle Swarm Optimisation) is used to produce a 'swarmic sketch' of the attended line.

Throughout the process of sketching, the swarms leave traces of themselves on the digital canvas which they would then revisit over and over again in an attempt to re-sketch themselves. This process is introduced in the context of autopoiesis.

Having associated the rendering process with the concepts of attention and autopoiesis, the performance of the participating swarms creates a unique, non-identical sketch each time the 'artist' swarms embark on interpreting the input line drawings. The complexity of the initial drawing reduces after each cycle and the *Swarmic Autopoietic System* continues this process until the digital canvas reaches its simplest form: emptiness. Additionally, a brief discussion is provided on the philosophical aspects of the autopoietic artist.

1 Introduction

Studies of the behaviour of social insects (e.g. ants and bees) and social animals (e.g. birds and fish) have proposed several new metaheuristics for use in collective intelligence. Natural examples of swarm intelligence that exhibit a form of social interaction are fish schooling, birds flocking, ant colonies in nesting and foraging, bacterial growth, animal herding, brood sorting etc.

Although producing artistic works through the use of swarm intelligence techniques have been previously explored, this work explores the concepts of attention and autopoiesis through this type of collective intelligence, which emerges through the interaction of simple agents (representing the social insects and animals) in nature-inspired algorithms, namely Stochastic Diffusion Search (SDS) [8, 3] and Particle Swarm Optimisation (PSO) [20].

In this work, SDS is deployed to enforce the idea of attention to area of the search space (digital canvas with line drawings) where there are more details (i.e. more lines or points); once the area of attention is identified, PSO through its particles, traces the points of the lines selected and its particles' movement are visualised on the canvas. As attention moves from one area of the original line drawing to another, a sketch is produced which is the collective result of the SDS-led attention and millions of simple interactions facilitated by PSO algorithm.

In the last couple of years, there has been several research work utilising the two aforementioned swarm intelligence algorithms; while scientific merits of integrating these algorithms are investigated in detailed (e.g. [6, 4]), their artistic capabilities have been detailed in several publications along with some philosophical arguments on the computational creativity of such systems (e.g. [1, 5, 2]).

In the next section a brief overview of some of the work in generative art and swarm intelligence is provided. Afterwards, the swarm intelligence algorithms used are explained. Subsequently, a historical perspective of attention is presented, followed by explanation on the attention and tracing mechanisms associated with the two swarm intelligence algorithms introduced in the paper, providing details on the performance of the computer-generated nature-inspired attentive swarms in re-interpreting the original line drawings. Subsequently, the concept of autopoietic swarmic artist is explained in the philosophical context of autopoiesis. Finally the paper is concluded and possible future research and suggested.

2 Generative Art and Swarm Intelligence

Among the many works in the field of generative art are research on swarm painting (e.g. [26, 7, 35, 36]), ant colony paintings (e.g. [15, 25, 30]) and other multi-agent systems (e.g. RenderBots [29] and the particle-based non-evolutionary approach of Loose and Sketchy Animation [13]).

In most of the swarm-based work mentioned above (e.g. [26, 7, 35, 36, 15]), the painting process does not re-work an initial drawing, but rather focuses on presenting "random artistic patterns", somewhere between order and chaos [36]. Other classes of research (e.g. by Schlechtweg et al. [29] and Curtis [13]) are based on reworking an initial drawing. There is a significant number of related papers in the area of non-photorealistic rendering; particularly, many papers approach drawing and painting using the optimisation framework (where optimisation and generative techniques are utilised an artistic context). Furthermore, particles have been used for stippling and other aesthetic styles in numerous papers. Turk and Bank's work [34] is an early example of optimising particle positions to control a stroke-based rendering. Hertzmann [16] optimised a global function over all strokes using a relaxation approach. In one of his works, Collomosse [12] used a global genetic algorithm to define a rendering algorithm. More recently, Zhao et al. [40] deployed an optimisation-

[1] Corresponding author

based approach to study the stroke placement problem in painterly rendering, and presented a solution named stroke processes, which enables intuitive and interactive customisation of painting styles.

3 Communication in Social Systems

Communication – social interaction or information exchange – observed in social insects and social animals plays a significant role in all swarm intelligence algorithms, including SDS and PSOs. Although in nature it is not only the syntactical information that is exchanged between the individuals but also semantic rules and beliefs about how to process this information [21], in typical swarm intelligence algorithms only the syntactical exchange of information is taken into account.

In the study of the interaction of social insects, two important elements are the individuals and the environment, which result in two integration schemes: the first is the way in which individuals self-interact (interact with each other) and the second is the interaction of the individuals with the environment [9]. Self-interaction between individuals is carried out through recruitment strategies and it has been demonstrated that, typically, various recruitment strategies are used by ants [17] and honey bees. These recruitment strategies are used to attract other members of the society to gather around one or more desired areas, either for foraging purposes or for moving to a new nest site.

In general, there are many different forms of recruitment strategies used by social insects; these may take the form of global or local strategies; one-to-one or one-to-many communication; and the deployment stochastic or deterministic mechanisms. The nature of information sharing varies in different environments and with different types of social insects. Sometimes the information exchange is quite complex where, for example it might carry data about the direction, suitability of the target and the distance; or sometimes the information sharing is simply a stimulation forcing a certain triggered action. What all these recruitment and information exchange strategies have in common is distributing useful information throughout their community [14].

However, in many hive-based (flock-based) agents – similar to the ones deployed in this work – the benefits of memory and communication seem obvious, but as argued in [28], these abilities are not beneficial in every environment, depending on the way resources are clustered throughout the environment and whether the quality of the food sources is sufficiently high.

The algorithms reported in this paper both rely on memory and communication to enable the agents explore various parts of the search space; albeit the communication methods outlined herein are less greedy than the one presented in [28]. Furthermore, the particular effect communication has on the "creative" act of the swarm-based algorithms used in this work is under further investigation.

The parable of 'The Blind Men and the Elephant' suggests how social interactions can lead to more intelligent behaviour. This famous tale, set in verse by John Godfrey Saxe [27] in the 19th century, characterises six blind men approaching an elephant. They end up having six different ideas about the elephant, as each person has experienced only one aspect of the elephant's body: wall (elephant's side), spear (tusk), snake (trunk), tree (knee), fan (ear) and rope (tail). The moral of the story is to show how people build their beliefs by drawing them from incomplete information, derived from incomplete knowledge about the world [21]. If the blind men had been communicating about what they were experiencing, they would have possibly come up with the conclusion that they were exploring the heterogeneous qualities that make up an elephant.

4 Attention & Creativity in the Swarms

In this section, the attention mechanism, which is controlled by the SDS algorithm is detailed. This is followed by the process through which the PSO algorithm utilises the output of the SDS-led attention to visualise the particles movements on the digital canvas which produces the final sketch rendered by the swarms. Details of the aforementioned swarm intelligence techniques are provided in the Appendices A and B; further details about attention along with some definitions are provided in Appendix C.

4.1 Attention Mechanism

The input digital image consists of line drawings (Fig. 1) where each line is formed up of a series of points (the image on the left is after one of Matisse sketches).

The swarms's attention in this work is controlled by the level of intensity of drawings within a specific radius, ra of an agent. In other words, the intensity or fitness of an agent, $f_{i,(x,y)}$, where i is the agent number and (x, y) is the coordinate of the agent in the search space (input image), is calculated by the number of points constituting the drawing within the radius ra (see Fig. 2a).

As mentioned earlier in Section A, each agent has two components: status, which is a boolean value and hypothesis. The hypothesis of each agent in this work is the (x, y) coordinate which is used to calculate the fitness, $f_{i,(x,y)}$, of the agents located at any particular pixel within the input image.

After the agents are randomly initialised throughout the search space (Fig. 3a), in order to determine the status of an agent, i, within the swarm (test phase), its fitness, f_i, is calculated as explained above and another agent, r, is randomly selected; if $f_i > f_r$ (i.e. the agent i is located in a more line intense area), agent i is set active, otherwise inactive (Fig. 3b illustrates active agents in red and inactive ones in black).

In the diffusion phase, as in standard SDS, each inactive agent randomly picks another one. If the randomly selected agent is active, the inactive agent adopts the hypothesis of the active one. However, if the selected agent is inactive, the selecting agent generates a random hypothesis (x, y) from the search space.

After n number of test-diffusion phases cycles, the biggest cluster of the agents is identified and the closest point (p_c) to the cluster is calculated. Once the (x, y) coordinate of the point is retrieved, the starting and end points of the line is extracted and a string of (x, y) coordinates from starting to end point of the line is passed on to the PSO particles to trace one by one. Fig. 2b shows that when a point is selected, the two ends of the line (i.e. starting and end points) are identified.

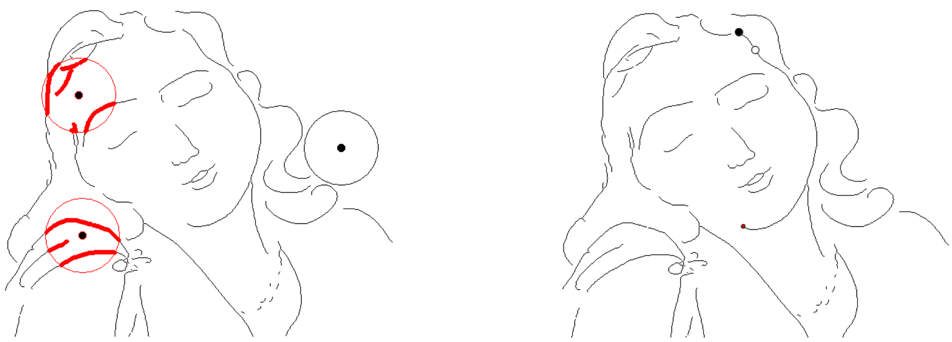
4.2 Tracing Mechanism

The points constituting the lines of the line drawing are treated as targets by the PSO algorithm. Thus, the particles aim to trace these points one at a time until reaching the end of the line (the algorithm tries to minimise the distance between the particles' positions and the points it aim to track).

Particle's movement is visualised on the canvas (i.e. trajectory of the particles moving from position (x_0, y_0) to (x_1, y_1) and so forth). The adopted PSO algorithm is briefly presented in Appendix B (more

Figure 1. Input images: series of points forming line drawings.

Figure 2. (a) Agent's fitness: in this figure, the (x, y) coordinates of three exemplar agents are illustrated with black dots in the centre of the circles; the highlighted points of the line drawing within each circle contribute towards the fitness of the agent, $f_{i,(x,y)}$. (b) Selected line: The hollow circle represents the selected point, p_c and the two ends of the line – start and end – where p_c resides, are highlighted in black and red, respectively.



(a) (b)

technical details on the behaviour of particles are reported in a previous publication [1]).

Input to PSO algorithm is a series of points forming up a line (whose starting and end points, as well as the closest point to the agents' biggest cluster, p_c, are extracted as mentioned above). The algorithm is then instructed to trace the line commencing from p_c to the beginning of the line, and them back towards the end of the line drawing. Once the line is traced, it is removed from the search space and the other lines are considered one by one according to the attention mechanism deployed.

This process ensures that in addition to the potential aesthetic of the swarms' final sketches, the process of sketching is enriched with attention to details.

5 Autopoiesis and Autopoietic Artist

In 1972, the Chilean biologists, Humberto Maturana and Francisco Varela, coined the term autopoiesis, referring to the self-maintaining chemistry of living cells [23]. Autopoiesis is composed of two Greek words, 'auto', meaning self, and 'poiesis' meaning production or creation.

There are many ways to think about systems that create products we socially conceive of as art. This paper is inspired by Alfred North Whitehead's *process view of organisation*, viewed though the transformational conceptual-lens of *autopoietic theory* (Maturana and Varela [38]); according to which we view a creative system as a clearly delineated and identifiable network of *continuously operational* component producing processes and concomitant elements, bounded as an autonomous entity *within its own artistic environment*.

The continual *creative swarmic processes* of our autopoietic artists' attention and reconstitution (sketching) mechanisms are detailed sections C, 4 of this paper and are illustrated in accompanying video, which displays her behaviour as she *artistically decodes* a line-sketch of an abstract painting by Willem De Kooning[2].

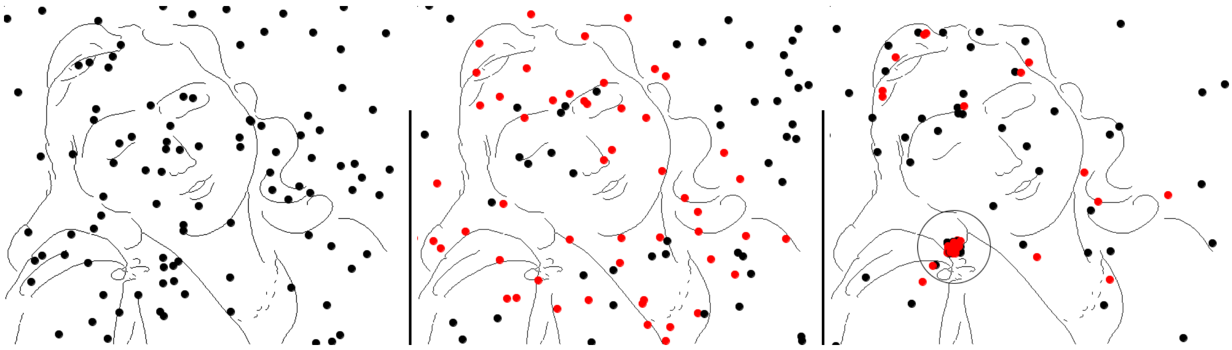
As observed in section 4, the 'autopoietic' artist is composed of two functionally distinct types of agent: (i) a swarm of *attending* agents, akin to ants (and governed by the principles of Stochastic Diffusion Search) and (ii) a swarm of *drawing* agents akin to birds (and governed by the principles of a Particle Swarm Optimiser). The job of the attending agents is to select areas of *meaning*[3] for the drawing agents to 're-interpret'.

Our 'autopoietic' artist is thus continually engaged in a process of sensing her environment and reconstituting it (by iteratively first

[2] In our case the artistic environment is initially a sketch of Kooning's abstract canvas, displayed initially in the right-hand panel of the video; with the creative output, initially a tabula-rasa, displayed on the left.

[3] For example, in our system we have defined such an area of interest (or 'meaning distinction') to be a line situated in a *complex* region of the image; an area that is rich/dense in comparison with other lines. Thus, by suitably redefining the distinction deployed by the population of Stochastic Diffusion agents (as described in Section V), we can modify what constitutes 'meaning' for the autopoietic artist as she interacts with her creative context/environment.

Figure 3. SDS stages: (a) Initialisation; (b) Test Phase; (c) Diffusion



(a) (b) (c)

choosing a line in the scene and re-rendering it). The bounds of the autopoietic artist are defined by the shifting movements of the swarms that comprise her; the elements of the autopoietic artist are the agents of the swarms; the behaviour of each swarm is fully defined by the behaviour of its agents (SDS and PSO); the bounds of the swarms are defined by the hypotheses (positions) of all the SDS agents, whose behaviour changes and in turn modifies the bounds; the components of the boundaries are produced by the interactions of the components of the unity, by transformation of previously produced hypotheses; and because the iterative re-initialisation of the SDS agent-hypotheses are produced by the interactions of the SDS swarm (and all other PSO agents participate as necessary permanent constitutive components in the production of other components), Varela et al's criteria [37] for an autopoietic entity are appropriately instantiated in the organisation of our 'autopoietic' artist *in the creative space in which her creative unity exists*.

Thus, following Luhmann's conception of *information processing* [22], we view the working autopoietic artist as entailing a reduction in complexity, ravenously consuming 'meaning-distinctions' within her environment; in this way the autopoietic artist iteratively decodes her environment (the De Kooning abstract) by continuously first selecting, then processing, areas of *meaning*

Over time, with her artistic 'interest' drawn to areas of rich complexity, the autopoietic artist, so construed, iteratively erases meaningful-distinctions (lines) in her current artistic context, so gradually simplifying the structure of the work. By iteratively focussing on meaning-distinctions *as-areas-of-rich-complexity*, as the decoding process unfolds it sometimes leads to a less complex (line) structure and ultimately may result in an empty canvas; therein reifying the artwork's 'death' and the tabula rasa (see Fig. 4a)[4].

Alternatively, by refocussing the autopoietic artist's reflections on 'meaning' (as explored by the Stochastic Diffusion swarm) onto different constitutive elements, and modifying her reconstitution (of the resulting artistic structure), different behaviours of autopoietic creativity can be induced. E.g. By insisting that the reconstitutive processes must generate as many elements of 'meaning-distinction' as they consume, the induced autopoietic processes becomes less likely to fade away and more open-ended in their creative endeavour (see Fig. 4 b)[5].

[4] Link to the video of Gluttonous Swarms: http://doc.gold.ac.uk/~map01mm/Swarmic_Sketches/deKooning_GluttonousSwarms.mov

[5] Link to the video of Contented Swarms: http://doc.gold.ac.uk/~map01mm/Swarmic_Sketches/deKooning_ContentedSwarms.mov

6 Conclusion

This works deploys two swarm intelligence algorithms: Stochastic Diffusion Search (mimicking the behaviour of ants foraging) and Particle Swarm Optimisation (mimicking the behaviour of birds flocking). The former is utilised for facilitating the attention mechanism and the latter is used for regulating the swarmic sketching process. In other words, swarms of ants and birds set off to decode a complex painting by Willem De Kooning in their own swarmic way. The step-by-step behaviour of the swarms, through the attention and tracing mechanisms is detailed.

The concept of attention is discussed in the context of the creativity and then the paper focuses on encapsulating the concept of autopoiesis in the behaviour of the the autopoietic artist described.

REFERENCES

[1] Mohammad Majid al-Rifaie, Ahmed Aber, and Mark Bishop, 'Cooperation of nature and physiologically inspired mechanisms in visualisation', in *Biologically-Inspired Computing for the Arts: Scientific Data through Graphics*, ed., Anna Ursyn. IGI Global, United States, (2012). ISBN13: 9781466609426, ISBN10: 1466609427.

[2] Mohammad Majid al-Rifaie and Mark Bishop, 'Weak vs. strong computational creativity', in *AISB 2012: Computing and Philosophy*, University of Birmingham, Birmingham, U.K., (2012).

[3] Mohammad Majid al-Rifaie and Mark Bishop, 'Stochastic diffusion search review', in *Paladyn, Journal of Behavioral Robotics*, volume 4(3), 155–173, Paladyn, Journal of Behavioral Robotics, (2013).

[4] Mohammad Majid al-Rifaie, Mark Bishop, and Tim Blackwell, 'Information sharing impact of stochastic diffusion search on differential evolution algorithm', in *J. Memetic Computing*, volume 4, 327–338, Springer-Verlag, (2012).

[5] Mohammad Majid al-Rifaie, Mark Bishop, and Suzanne Caines, 'Creativity and autonomy in swarm intelligence systems', in *J. Cognitive Computation*, volume 4, 320–331, Springer-Verlag, (2012).

[6] Mohammad Majid al-Rifaie, Mark John Bishop, and Tim Blackwell, 'An investigation into the merger of stochastic diffusion search and particle swarm optimisation', in *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, GECCO '11, pp. 37–44, New York, NY, USA, (2011). ACM.

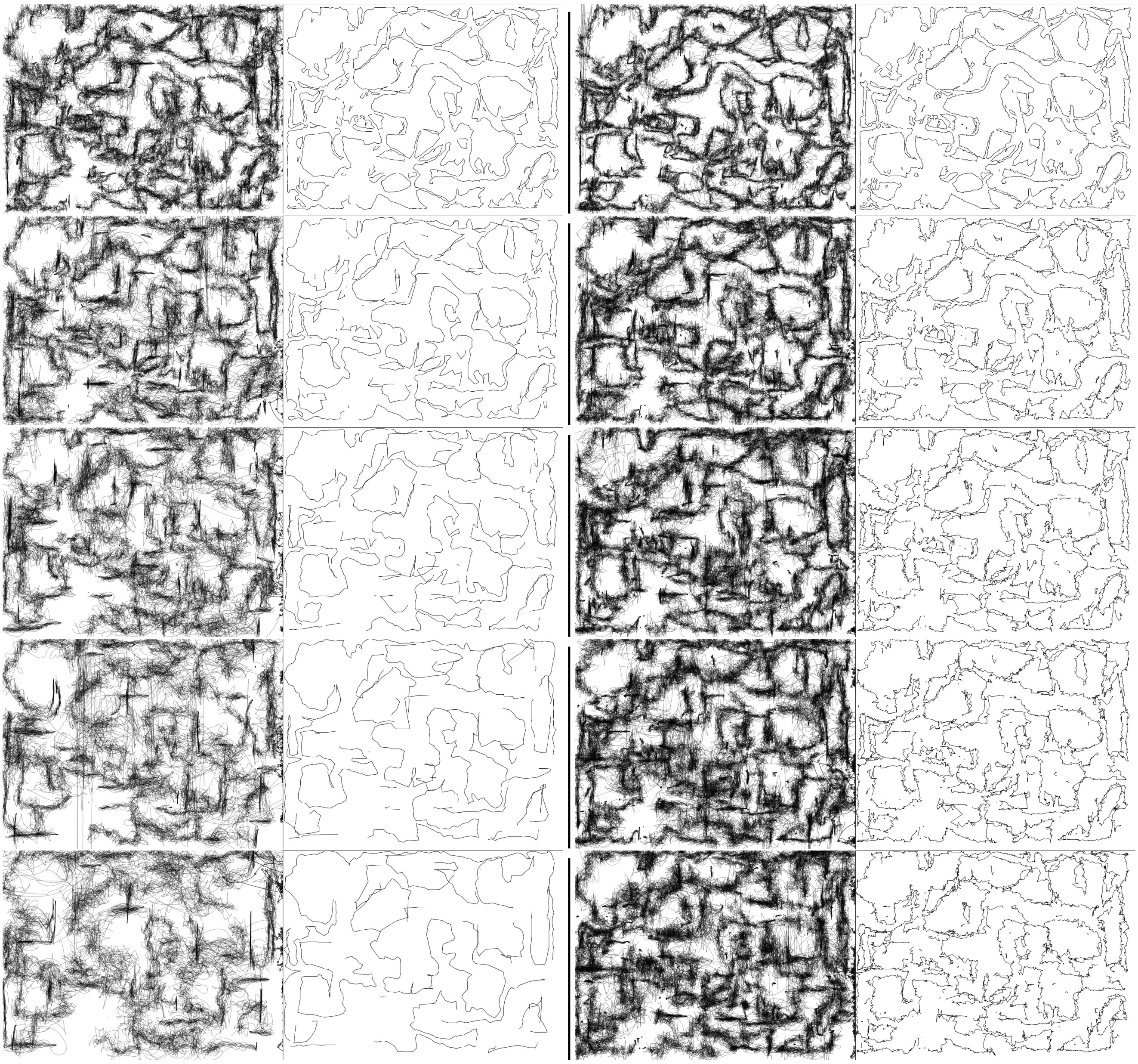
[7] S. Aupetit, V. Bordeau, N. Monmarche, M. Slimane, and G. Venturini, 'Interactive evolution of ant paintings', in *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*, volume 2, pp. 1376–1383, (2004).

[8] J.M. Bishop, 'Stochastic searching networks', in *Proc. 1st IEE Conf. on Artificial Neural Networks*, pp. 329–331, London, UK, (1989).

[9] E. Bonabeau, M. Dorigo, and G. Theraulaz, 'Inspiration for optimization from social insect behaviour', *Nature*, 406, 3942, (2000).

[10] D. Bratton and J. Kennedy, 'Defining a standard for particle swarm optimization', in *Proc of the Swarm Intelligence Symposium*, pp. 120–127, Honolulu, Hawaii, USA, (2007). IEEE.

Figure 4. Contented and Gluttonous Autopoietic Artist



(a) Gluttonous swarms (b) Contented swarms

[11] M. Clerc and J. Kennedy, 'The particle swarm-explosion, stability, and convergence in amultidimensional complex space', *Evolutionary Computation, IEEE Transactions on*, 6(1), 58–73, (2002).

[12] J. Collomosse and P. Hall, 'Genetic paint: A search for salient paintings', *Applications of Evolutionary Computing*, 437–447, (2005).

[13] C. J Curtis, 'Loose and sketchy animation', in *ACM SIGGRAPH 98 Electronic art and animation catalog*, p. 145, (1998).

[14] K. de Meyer, S.J. Nasuto, and J.M. Bishop, 'Stochastic diffusion optimisation: the application of partial function evaluation and stochastic recruitment in swarm intelligence optimisation', *Springer Verlag*, 2, Chapter 12 in Abraham, A. and Grosam, C. and Ramos, V. (eds), "Swarm intelligence and data mining", (2006).

[15] G. Greenfield, 'Evolutionary methods for ant colony paintings', *APPLICATIONS OF EVOLUTIONARY COMPUTING, PROCEEDINGS*, 3449, 478–487, (2005).

[16] A. Hertzmann, 'Paint by relaxation', in *Computer Graphics International 2001. Proceedings*, pp. 47–54. IEEE, (2001).

[17] B. Holldobler and E. O. Wilson, *The Ants*, Springer-Verlag, 1990.

[18] W. James, *The principles of psychology*, 1890.

[19] Henry Home Kames, *Elements of Criticism*, 1769.

[20] J. Kennedy and R. C. Eberhart, 'Particle swarm optimization', in *Proceedings of the IEEE International Conference on Neural Networks*, volume IV, pp. 1942–1948, Piscataway, NJ, (1995). IEEE Service Center.

[21] James F. Kennedy, Russell C. Eberhart, and Yuhui Shi, *Swarm intelligence*, Morgan Kaufmann Publishers, San Francisco ; London, 2001.

[22] Niklas Luhmann, *Social systems*, Stanford University Press, 1995.

[23] Humberto R Maturana and Francisco J Varela, *Autopoiesis and cognition: The realization of the living*, number 42, Boston Studies in the Philosophy of Science 42. Dordecht: D. Reidel Publishing Co., 1973. ISBN: 90-277-1015-5.

[24] M. Moglich, U. Maschwitz, and B. Holldobler, 'Tandem calling: A new

kind of signal in ant communication', *Science*, 186(4168), 1046–1047, (1974).
[25] N. Monmarche, S. Aupetit, V. Bordeau, M. Slimane, and G Venturini, 'Interactive evolution of ant paintings', in *2003 Congress on Evolutionary Computation*, ed., B. McKay et al, volume 2, pp. 1376–1383. IEEE Press, (2003).
[26] L. Moura and V. Ramos, 'Swarm paintings–nonhuman art', *ARCHITOPIA book, art, architecture and science*, 5–24, (2007).
[27] John Godfrey. Saxe, D.A. Lathen, and B. Chief, 'The Blind Man and the Elephant', *The Poems of John Godfrey Saxe*, (1882).
[28] P. Schermerhorn and M. Scheutz, 'The impact of communication and memory in hive-based foraging agents', in *Artificial Life, 2009. ALife'09. IEEE Symposium on*, pp. 29–36, (2009).
[29] S. Schlechtweg, T. Germer, and T. Strothotte, 'Renderbots–multi-agent systems for direct image generation', in *Computer Graphics Forum*, volume 24, pp. 137–148, (2005).
[30] Y. Semet, U. M O'Reilly, and F. Durand, 'An interactive artificial ant approach to non-photorealistic rendering', in *Genetic and Evolutionary Computation–GECCO 2004*, pp. 188–200, (2004).
[31] Y. Shi and R. C Eberhart, 'Parameter selection in particle swarm optimization', *Lecture notes in computer science*, 591–600, (1998).
[32] C. C. W. Taylor, 'Berkeley's theory of abstract ideas', *The Philosophical Quarterly*, 97115, (1978).
[33] E. B. Titchener, *Lectures on the elementary psychology of feeling and attention*, The Macmillan Company, 1908.
[34] G. Turk and D. Banks, 'Image-guided streamline placement', in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 453–460. ACM, (1996).
[35] P. Urbano, 'Playing in the pheromone playground: Experiences in swarm painting', *Applications on Evolutionary Computing*, 527–532, (2005).
[36] P. Urbano, 'Consensual paintings', *Applications of Evolutionary Computing*, 622–632, (2006).
[37] Francisco G Varela, Humberto R Maturana, and Ricardo Uribe, 'Autopoiesis: the organization of living systems, its characterization and a model', *Biosystems*, 5(4), 187–196, (1974).
[38] Francisco J Varela and Humberto R Maturana, 'De máquinas y seres vivos: Una teoría sobre la organización biológica', *Santiago de Chile: Editorial Universitaria*, (1973).
[39] Phuong Vu, *Historical Overview of Research on Attention*, Sage Publications, Incorporated, 2003.
[40] M. Zhao and S.C. Zhu, 'Customizing painterly rendering styles using stroke processes', in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, pp. 137–146. ACM, (2011).

Appendices

A Stochastic Diffusion Search

This section introduces Stochastic Diffusion Search (SDS) [8] – a swarm intelligence algorithm – whose performance is based on simple interaction of agents. This algorithm is inspired by one species of ants, *Leptothorax acervorum*, where a 'tandem calling' mechanism (one-to-one communication) is used, the forager ant that finds the food location recruits a single ant upon its return to the nest; therefore the location of the food is physically publicised [24].

The SDS algorithm commences a search or optimisation by initialising its population and then iterating through two phases (see Algorithm 1)

Algorithm 1 SDS Algorithm

```
Initialise agents
While (stopping condition is not met)
   For each agent
       Test hypothesis and determine activity
   For each agent
       Diffuse hypothesis
End While
```

In the test phase, SDS checks whether the agent hypothesis is successful or not by performing a hypothesis evaluation which returns a boolean value. Later in the iteration, contingent on the precise recruitment strategy employed (in the diffusion phase), successful hypotheses diffuse across the population and in this way information on potentially good solutions spreads throughout the entire population of agents. In other words, each agent recruits another agent for interaction and potential communication of hypothesis.

In standard SDS (which is used in this paper), *passive recruitment mode* is employed. In this mode, if the agent is inactive, a second agent is randomly selected for diffusion; if the second agent is active, its hypothesis is communicated (*diffused*) to the inactive one. Otherwise there is no flow of information between agents; instead a completely new hypothesis is generated for the first inactive agent at random. Therefore, recruitment is not the responsibility of the active agents.

B Particle Swarm Optimisation

A swarm in Particle Swarm Optimisation (PSO) algorithm comprises of a number of particles and each particle represents a point in a multi-dimensional problem space. Particles in the swarm explore the problem space searching for the optimal position, which is defined by a fitness function.

Each particle has a position x, a velocity v, and a memory, p, containing the best position found so far during the course of the optimisation, and this is called the personal best (pbest). p can also be thought of as a particle 'informer'. Particles participate in a social information sharing network. Each particle is informed by its neighbours within this network, and in particular, the best position so far found in the neighbourhood, is termed the neighbourhood best. The position of each particle is dependent on the particle's own experience and those of its neighbours.

The standard PSO algorithm defines the position of each particle by adding a velocity to the current position. Here is the equation for updating the velocity of each particle:

$$v_{id}^{t} = wv_{id}^{t-1} + c_1 r_1 \left(p_{id} - x_{id}^{t-1}\right) + c_2 r_2 \left(g_{id} - x_{id}^{t-1}\right) \tag{1}$$

$$x_{id}^{t} = v_{id}^{t} + x_{id}^{t-1} \tag{2}$$

where w is the inertia weight whose optimal value is problem dependent [31]; $\vec{v}_{id}^{t-1}$ is the velocity component of particle i in dimension d at time step $t-1$; $c_{1,2}$ are the learning factors (also referred to as acceleration constants) for personal best and neighbourhood best respectively (they are constant); $r_{1,2}$ are random numbers adding stochasticity to the algorithm and they are drawn from a uniform distribution on the unit interval $U(0,1)$; p_{id} is the personal best position of particle x_i in dimension d; and g_{id} is the neighbourhood best. Therefore, PSO optimisation is based on particles' individual experience and their social interaction with other particles. After position and velocity updates, the positions of the particles are evaluated and the memories p are updated, if a better position has been found.

In this paper, Clerc and Kennedy's PSO (PSO-CK [11]) or constriction PSO is used:

$$v_{id}^{t} = \chi \left(v_{id}^{t-1} + c_1' r_1 \left(p_{id} - x_{id}^{t-1}\right) + c_2' r_2 \left(g_{id} - x_{id}^{t-1}\right)\right) \tag{3}$$

where $\chi = 0.72984$ [10], which is reported to be working well in general, is used in this work.

C Attention

For centuries, attention has been preoccupying the minds of philosophers and psychologists and scientists. The concept of attention has been studied mostly in psychology and neuroscience (see Table 1.1 in Phuong Vu: Historical Overview of Research on Attention, in [39] for more details) and there has been considerably less notable interest on attention within the field of computational creativity.

In the early 18^{th} century attention was mostly seen as a way of abstraction (see Berkeley's 1710 theory of abstract ideas in [32]):

> "[It] must be acknowledged that a man may consider a figure merely as triangular, without attending to the particular qualities of the angles or relations of the sides. So far he may abstract, but this will never prove that he can frame an abstract general, inconsistent idea of a triangle. "

By 1769, when Henry Home Kames added the appendix of 'Terms Defined or Explained' to his *Elements of Criticism* [19], attention's role as a regulator of cognitive input was regarded as definitive of it:

> "Attention is that state of mind which prepares one to receive impressions. According to the degree of attention objects make a strong or weak impression. Attention is requisite even to the simple act of seeing."

Thus, regulating cognitive and sensory inputs was associated to attention. Later, William James in The Principles of Psychology in 1890 [18] offered a more comprehensive description of attention (i.e. focalisation, etc.):

> "Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence [...]" (p. 403-404)

and few pages further, he continues:

> "Each of us literally chooses, by his ways of attention to things, what sort of a universe he shall appear to himself to inhabit." (p. 424)

Two decades later, in 1908, as emphasised by E.B. Titchener [33], attention was given a greater significance :

> "What I mean by the 'discovery' of attention is the explicit formulation of the problem: the recognition of its separate status and fundamental importance; the realization that the doctrine of attention is the nerve of the whole psychological system, and that as men judge of it, so they shall be judged before the general tribunal of psychology."

and its importance grew over the years in psychology and neuroscience. Although the concept of attention might have been present in the work of some researchers in the field of computational creativity, the focus on attention has not been equally considerable among researchers in this field; perhaps, partly because there has not been a clear definition on attention.

The longer term value of creativity judgements in computational creativity

Anna Jordanous [1]

Abstract. During research to develop the Standardised Procedure for Evaluating Creative Systems (SPECS) methodology for evaluating the creativity of 'creative' systems, in 2011 an evaluation case study was carried out. The case study investigated how we can make a 'snapshot' decision, in a short space of time, on the creativity of systems in various domains. The systems to be evaluated were presented at the International Computational Creativity Conference in 2011. Evaluation was performed by people whose domain expertise ranges from expert to novice, depending on the system. The SPECS methodology was used for evaluation, and was compared to two other creativity evaluation methods (Ritchie's criteria and Colton's Creative Tripod) and to results from surveying people's opinion on the creativity of the systems under investigation. Here, we revisit those results, considering them in the context of what these systems have contributed to computational creativity development. Five years on, we now have data on how influential these systems were within computational creativity, and to what extent the work in these systems has influenced further developments in computational creativity research. This paper investigates whether the evaluations of creativity of these systems have been helpful in predicting which systems will be more influential in computational creativity (as measured by paper citations and further development within later computational systems). While a direct correlation between evaluative results and longer term impact is not discovered (and perhaps too simplistic an aim, given the factors at play in determining research impact), some interesting alignments are noted between the 2011 results and the impact of papers five years on.

1 Introduction

In [8], the Standardised Procedure for Evaluating Creative Systems (SPECS) methodology was developed as a tool for evaluating the creativity of software developed within computational creativity research. SPECS is summarised in Table 1. As part of the research to develop SPECS, two case studies were carried out; this paper focuses on the second case study reported in [8].

The case study we focus on here was carried out at the 2011 International Computational Creativity Conference (ICCC'11), and explored to what extent creativity evaluation methods can be applied across creative systems demonstrating different types of creativity rather than focusing exclusively on systems operating specifically within one creative domain. This case study specifically explored the scenarios where we do not have the full information desired for evaluation, or where we may have only limited time to complete evaluation, or be limited as to who can perform evaluation. This was motivated by the issue that we often wish to compare one system's creativity against that of others, but for various reasons may not have the full information available to us that we would like, or may be working under time pressures. Section 1.1 discusses this.

Four different evaluation methods were applied to evaluate the creativity of five systems: the collage generation module for the artistic system *The Painting Fool* [2] [4] [4]; a poetry generator [17]; the *DARCI* system [15] for generating images to illustrate given adjectives; a reconstruction of the *MINSTREL* story-telling system [23] [22]; and a musical soundtrack generator matching emotions in a narrative to the music generated [12]. The evaluation methods used in this case study were: SPECS [8, 7]; Ritchie's empirical criteria [18]; Colton's creative tripod [2]; and asking people's opinion on how creative each system was. In each evaluation, the judges worked with limited information and time.

The 2011 case study resulted in formative evaluative feedback for the systems to help researchers develop the creativity of their system. Section 2 summarises the generated feedback, which is fairly detailed even given the time and information pressures.[2] As the creative domain varies across systems, comparisons between systems became less relevant: the systems were designed to perform different tasks, requiring different interpretations of creativity [16, 8]. Hence the focus in this case study was on evaluating individual systems - though some interesting comparisons could made between systems where there are commonalities in creative priorities of that domain.

Looking back at this case study five years later, we can now see what contributions each system has made to the development of computational creativity research over the past five years, as measured in citations each 2011 paper has received, and in tracing what development work has been done since 2011 that can be directly related to the 2011 systems. This is a beneficial exercise: given that value is an important part of creativity, we could hypothesise that those systems judged more creative can have had more value to the computational creativity community over the last few years. Hence we can test on our (small)[3] sample as to whether initial judgements of the creativity of these systems give us information as to which systems will provide greater contributions to computational creativity research.

1.1 Digital resource availability for evaluation

Creative systems are by their nature likely to be different to every other system and it is useful to see how a creative task has been approached in different ways, when we are investigating that task

[1] School of Computing, University of Kent, Medway Building, Chatham Maritime, Kent, UK, email: a.k.jordanous@kent.ac.uk

[2] An unexpected but beneficial extra finding of the evaluation was that it highlighted which ICCC'11 presentations had contained adequate information for judging the creativity of their systems.

[3] Sample size means that results are indicative rather than conclusive.

Table 1. The Standardised Procedure for Evaluating Creative Systems (in summarised form)

1. Identify a definition of creativity that your system should satisfy to be considered creative
2. Using Step 1, clearly state what standards you use to evaluate the creativity of your system.
3. Test your creative system against the standards stated in Step 2 and report the results.

comptationally. There may be systems that are related in some way to systems that we are developing, where it would be of interest to learn more about the research behind that system(s); in particular it would be useful to gain knowledge from seeing the system in operation, as well as reading published reports. As example, in evaluating the GAmprovising musical improvisation system against GenJam and Voyager [8], several useful insights arose for the development of GAmprovising from evaluating GenJam and Voyager.

It is more straightforward to evaluate systems for which we have full access to view and run the source code, with as much time available as we need, all necessary data and a line of communication with the system developers. This ideal evaluation scenario, however, is often not possible.

Taking time constraints first, the amount of time researchers can spend on evaluation is partly dictated by factors such as the allocation of researchers' time (particularly when conducting multiple projects or when time allocations are dictated by funding), deadlines for conferences etc., time demands within a project and the scheduling of other tasks to be completed within the project. Further demands on researchers' time include teaching, administration and other research work. There are often also constraints on the time and availability at appropriate times of other people involved in performing the evaluation. Another important issue impacting upon evaluation is if there are problems with availability of relevant software, data or more detailed information for a creative system(s) that we are interested in.

We could choose not to use systems for comparative evaluation if we do not have the full access and data that we would like; however, while this reduces the evaluation workload, it also sacrifices the opportunity to learn from this system. Alternatively, we can include systems in comparative evaluation even if we only have partial information for that system, keeping aware of the constraints on what we can learn from such evaluation but taking advantage of what is available, for formative feedback into the development of our own existing and future systems. Without evaluation of other systems:

> 'lessons from the past are difficult to learn' [1, p. 149, reflecting on the lack of availability of computer artworks and their related research resources]
> 'without cultural artifacts, civilization has no memory and no mechanism to learn from its successes and failures. And paradoxically, with the explosion of the internet, we live in what Danny Hillis has referred to as our "digital dark age".'[4]

When would we wish to learn from other existing systems? Systems of historical interest would have intrinsic value, even if the system can no longer be obtained. For example, James Meehan's TALESPIN system [11] has proven to be a seminal work in the field of story generation, even though the code has been lost and only a 'micro-TALESPIN' version exists today [10], which was itself published over 30 years ago. Similarly, Christopher Longuet-Higgins produced software for expressive musical performance which was widely praised by those who heard it [5, and personal communications with: Darwin, 2012; Dienes, 2012; Torrance, 2012; Thornton, 2012]. Unfortunately, the system was not made available as code or in a published report before Longuet-Higgins' death, and the code was archived but now cannot be restored due to the use of obsolete data storage formats.[5] We can learn from what our peers are doing in closely related research areas, and also by cross-applying work from less related areas to our own interests.

As Robey [19] has remarked, research that produces computer programs is surrounded by issues of software sustainability. Unfortunately, even for more recent systems, it can be difficult to retrieve all information necessary for full evaluation of a system. Bentkowska-Kafel [1] and Robey [19] have highlighted the speed at which current or cutting-edge digital resources can quickly become obsolete or lost, sometimes in a matter of only a few years.

> 'digital information lasts forever - or five years, whichever comes first.' [20, p. 2]

Jordanous [8] discusses several potential reasons:[6]

- Digital resources such as source code may not have been made available publicly.
- The researchers may not be available to contact (they may have left academia, or passed away) or may have moved onto other projects and forgotten details of the project of evaluative interest.
- Code may be unavailable or obsolete even if obtained. [6, pp. 34-35] identifies various reasons why available code may become unusable, including hardware or software obsolescence, third party dependencies, proprietary or poorly documented code as well as concerns about protecting intellectual property rights (especially in more competitive scenarios).
- Published code/digital resorces may not remain available long-term, for example if funding runs out for online hosting costs.

1.2 Selection of the creative systems being evaluated

The International Computational Creativity Conference (ICCC) is an annual international conference series dedicated to computational creativity research. Since its inception in 2010 it has been the main presentation venue for the latest findings in computational creativity research, taking over this role from the previous International Joint Workshops in Computational Creativity (IJWCC), from which the conference series evolved. ICCC'11 was held in Mexico in April 2011. Many creative systems were presented, demonstrating various types of creativity in different domains.

At ICCC'11, papers were presented to the conference audience in talks lasting seven minutes (a particularly brief amount of time for talks). There is a limit to what can be presented in this time and

[4] Original source unattributed, quote taken from $http://archive.org/about$ (last accessed January 2016).

[5] According to personal communications with Jeremy Maris and other IT support staff at the University of Sussex, where Longuet-Higgins' computer files were archived, and with a digital archive specialist, Gareth Knight.

[6] Jordanous [8] also discusses several recent efforts to promote software sustainability.

Table 2. The five systems from ICCC'11 that were evaluated for the 2011 case study.

Paper	System (if named)	Domain	Purpose
[4]	Module of *The Painting Fool*	Art	Collage generation
[17]	Adapted from an earlier system: McGonagall [9]	Poetry	Poetry generation
[15]	*DARCI*	Art	Image selection
[22]	Reconstruction of *MINSTREL*	Narrative	Story generation
[12]	-	Music	Soundtrack generation

it is unlikely that all desired information can be provided, but the ICCC'11 organisers posited that enough information could be delivered for the audience to become acquainted with the paper content. During the seven-minute talks at ICCC'11, presenters aimed to convey enough information about their paper so that people could discuss issues raised, in a group of related talks.

Five of the creative systems presented at ICCC'11 were selected for the 2011 case study, representing a variety of creative domains. These five systems were evaluated by two judges on their creativity using SPECS [8, 7, See Table 1 for a summary], based on the information presented in the seven-minute talks. The evaluation also generated qualitative feedback for the presenters of the evaluated systems, from two perspectives: the perceived creativeness of their system and the quality of information that they presented about their system in the brief time frame permitted. (The original purpose of this case study in 2011 was to test out the SPECS methodology for evaluating creative systems.)

One more point to note is that the systems evaluated in this case study were from a variety of domains, rather than just one domain. Some comparisons can be made between systems from different domains, and some interesting insight can be gained from doing so. On the whole, though, this paper acknowledges that such comparisons are less useful than comparisons made between systems from similar domains, as there are fewer areas of crossover so therefore less relevant information is available from the comparison. Some non-obvious conclusions may still however be reached this way, through viewing the systems from different perspectives. Comparing systems across different creative domains can also give a general (if limited) impression of relative progress in each domain.

Later, the information collected in SPECS evaluation was re-applied to use Colton's Creative Tripod as the evaluative method. Then analysis was carried out on the five selected systems using Ritchie's empirical criteria [18], and through asking people their opinion on systems' creativity. To replicate the time pressures and limits on available information in the latter stage of evaluation, judges were given only the information available in the paper, and had only a short amount of time to read the paper before evaluation.[7 8]

Of the five presentations in the first session (entitled 'The Applied'), the judges decided that three presented details of a computational creativity system that could be evaluated. Two further systems from the third session ('The Narrative') were also evaluated, for a total of five systems evaluated in this case study. These five systems, along with the papers they were presented in, the authors and the creative domain which the system operates in, are listed in Table 2.

[7] It is acknowledged that a closer replication could have been achieved; however the principles behind the evaluation remain the same (time and information availability pressures) and in this paper the emphasis is not on comparing the different evaluation methods, but on learning from their collective findings.

[8] Full details of the methodologies and how they were applied in this case study can be found in [8].

Examples of some of these systems' output are given in Figures 1 [4] and 2(a) [15], and in this example limerick poem from [17]:

> 'There **was** the young **la**dy called **Bright**.
> They could **tra**vel much **fas**ter than **light**.
> They **re**latively **left** one **day**. It sur**vived**.
> They **left**. On the **night**, she re**turned**.'
> [17, p. 9] *Where syllables are in bold, that syllable should be stressed when reading aloud the limerick.*

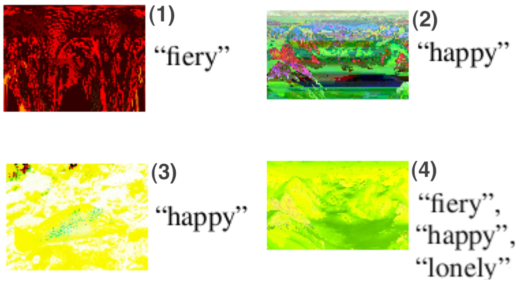
Figure 1. Collage generated by Cook & Colton's collage generation module in *The Painting Fool* system, on the theme of the current war in Afghanistan (Cook & Colton, 2011, Fig. 1, p. 2).

2 Results of the 2011 case study

2.1 Applying the SPECS methodology

The 14 components of creativity identified in [7] (see Figure 3) were used as a definition of creativity in a general context.

The judges recorded what general creative domain each system was designed to operate in (e.g. art, narrative generation). They also assessed their level of expertise and competence in that domain as either *Basic*, *Reasonable* or *Expert*. For each component, judges rated how successfully the system performed on that component requirements, giving a score between 0 (lowest) and 10 (highest). The judge rated the system based on the information given in the conference



(a) Output images, each intended to illustrate the adjective(s) listed to the right of that image (Norton et. al., 2011, Figs. 4-7, p. 14).

(a) Image A (b) Image B (c) Image C

(b) Inspiring set images (Norton et. al., 2011, Fig. 2, p. 13).

Figure 2. *DARCI* output, and the inspiring set of source images used to generate this output.

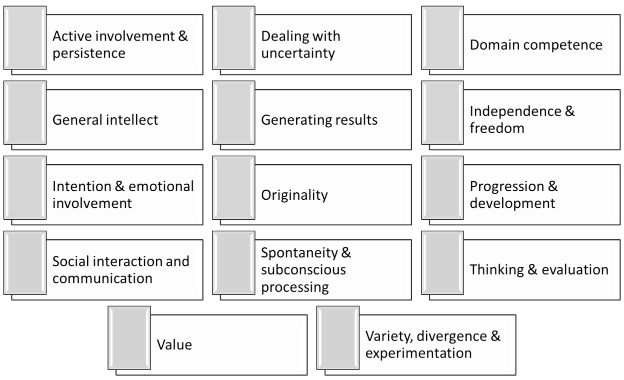


Figure 3. Jordanous's 14 components of creativity [7], derived through empirical analysis of the words used in texts about creativity.

talk; if they felt that not enough information was given about a particular component to provide a rating, then this rating was left blank. Each component was categorised according to how important the judge felt that component was for creativity in the domain which that system operated in. The contribution of that component to creativity in the system's domain was categorised according to how important that component was for creativity.

Jordanous [8] presents full results of what was learned from this case study; here our primary focus is in seeing how the creativity of each system was judged, relative to the other systems. To summarise:

- The collage generator [4] performed well at creating results, demonstration of intention and social abilities, but could improve its originality, value and ability to be spontaneous.
- The poetry generator [17] was good at creating results in a domain-competent way but needs to improve its ability to experiment and diverge and to a lesser extent could improve upon its originality, value and spontaneity.
- *DARCI* [15] showed strengths in social interaction, spontaneity, self-evaluation and production of results, but could perform better on originality and value.
- The story generator's [22] abilities to be original and to produce results were praised, though it could improve upon its inherent value, its ability to progress and develop and to work independently.
- The soundtrack generator [12] was considered valuable and competent in its domain, but could improve its ability to diverge and experiment.

Some systems performed better in evaluation, notably *DARCI*:

- *DARCI* [15] was rated highly on 50% of the components key to creativity in its domain, with the remaining systems scoring between 25% [4] and 33% [17, 12].
- Accounting for middling ratings as well, again *DARCI* was ahead of the other systems, with 75% of its key components receiving a high or middling rating. Three systems had 50-54% of its key components receiving high or middling ratings [17, 22, 12]. *The Painting Fool*'s collage generator [4] only received high or middling ratings for 25% of its key components.
- The reconstruction of *MINSTREL* [22] was the only system to receive a low rating for one of its key components, though it did have the largest number of key components to address.
- Quantifying the ranking data obtained such that high ratings score 2 points, middling ratings score 1 point and low ratings or unrated components score 0 points, with the total divided by the number of components considered key to that type of creativity by the judges,[9] overall rankings can be generated:

1. *DARCI*: 5/4 = 1.25 points.
2. *MINSTREL* reconstruction: 6/7 = 0.857 points (to 3 s.f.).
3. by Rahman & Manurung: 5/6 = 0.833 points (to 3 s.f.) and by Monteith et al.: 5/6 = 0.833 points (to 3 s.f.).
4. by Cook & Colton: 2/4 = 0.5 points.

2.2 Applying Ritchie's criteria

Ritchie's criteria were applied in a similar fashion to the applications reported in [18], except that (because of the Boolean way they are defined by Ritchie) the criteria were treated as a set of criteria which can be either satisfied or not satisfied, depending on whether a threshold value is reached or not. This approach better fits the As discussed in [8], Ritchie's criteria [18] concentrate almost exclusivelyon observations about the output of the system, measuring how typical that output is of the domain and how valuable the output is in the domain. (The criteria also include information on the *inspiring set* of input examples a system may have been constructed from.) An approach similar to that used in SPECS was adopted to meet these demands, with two judges asked to provide ratings. If the authors of the five of the 2011 case study systems had all provided examples of their systems (or links to examples) in their papers, then these could be used for evaluation using Ritchie's criteria. Ideally, details of inspiring sets would also be available for Ritchie's criteria to be fully applied. Unfortunately this was not always the case. Jordanous [8] discusses reasons for variability in available information, and reports efforts to locate additional examples of output and inspiring sets. It is

[9] This is of course one of several ways to quantify this information.

worth repeating here, however, that the point of this evaluation case study was to evaluate systems based on the information available, and work with incomplete information.

What can be done is to evaluate the results of the systems as and when presented in the papers, with no evaluations being performed for the Tearse et al. [22] and Monteith et al. [12] systems using Ritchie's criteria as these authors did not provide examples either in their papers or in online supplementary resources. Full details of the criteria calculations for each system are given in [8]. To summarise:

- For the collage generator, of 8 applicable criteria, 1 evaluated as TRUE (Criterion 10a) and 7 as FALSE (Criteria 1-4, 6, 7, 9).
- For the poetry generator, of 7 applicable criteria, 1 evaluated as TRUE (Criterion 10a) and 6 as FALSE (Criteria 1-4, 6, 9).
- For *DARCI*, of 10 applicable criteria, 2 evaluated as TRUE (Criteria 5, 10a) and 8 as FALSE (Criteria 1-4, 6-9).
- Neither the story generator or the soundtrack generator could be evaluated due to lack of information on their inspiring sets.

DARCI [15] was the only system for which two criteria (5, 10a) rather than one (10a) were true. It also had the fewest inapplicable criteria; the only inapplicable criteria were Criteria 11-18 which, it was noted earlier, could not be applied for any of these systems if the results set did not include items from the inspiring set.

The two criteria that *DARCI* satisfied were:

5. A decent proportion of the output are both suitably typical and highly valued.

10a. Much of the output of the system is not in the inspiring set, so is novel to the system.

The two other evaluated systems [4, 17] also satisfied the second of these criteria, 10a.

It is unclear in [18] how the criteria results should be analysed. Is *DARCI* [15] the most creative because it satisfied 2 criteria as opposed to 1, or is Rahman & Manurung's poetry generator [17] poetry generator most creative because it had least false criteria (6 as opposed to 7 for Cook & Colton [4] and 8 for Norton et al. [15])? Or should the number of inapplicable criteria be taken into account? It was decided that for this analysis, the percentage of applicable criteria that were true would be calculated for each system and this would be used to compare the systems' creativity. Therefore if a criterion is not applicable to a system, it is not considered for that system.

- Cook & Colton's collage generator [4] satisfied 1 out of 8 applicable criteria (12.5%).
- Rahman & Manurung's poetry generator [17] satisfied 1 out of 7 applicable criteria (14.3%).
- Norton et al.'s image generator [15] satisfied 2 out of 10 applicable criteria (20%).

These results place the *DARCI* image generator by Norton et al. [15] as the most creative system of the three, followed by Rahman & Manurung's poetry generator [17] and then Cook & Colton's collage generation module for *The Painting Fool* system [4]. For all three systems, though, only a small percentage of criteria were satisfied.

2.3 Applying Colton's creative tripod

In applying the creative tripod [2] for evaluation, we see that Colton's tripod qualities map to three of the 14 components used for SPECS:

- Skill ≈ *Domain Competence*.
- Imagination ≈ *Variety, Divergence and Experimentation*.
- Appreciation ≈ *Thinking and Evaluation*.

The evaluation data gathered on these three components could therefore be used to evaluate the systems using the creative tripod.

- The collage generator showed average imaginative abilities and there was a lack of information on other qualities, with mean ratings out of 10 of 5.0 for imagination but no data for skill or appreciation.
- The poetry generator demonstrated very good skilfulness and appreciation with average imagination, with mean ratings out of 10 of 8.5 for skill, 5.0 for imagination and 8.0 for appreciation.
- *DARCI* showed average to good all-round performance on the tripod qualities, with mean ratings out of 10 of 7.0 for skill, 7.5 for imagination and 6.0 for appreciation.
- The story generator performed reasonably well on all three tripod qualities although could improve its imaginative abilities, with mean ratings out of 10 of 8.0 for skill, 5.5 for imagination and 7.0 for appreciation.
- The soundtrack generator gave partial information on the tripod qualities, demonstrating average skill and imagination, with mean ratings out of 10 of 6.0 for skill, 5.0 for imagination but no data for appreciation.

Three systems emerge from this evaluation as 'balanced', i.e. with all three 'legs' present [17, 15, 22]. Monteith et al.'s system [12] could not be evaluated on its appreciative abilities and Cook & Colton's system [4] presentation lacked information on both its skill and appreciation. Both systems received only middling ratings in all cases where ratings were provided, apart from a 7/10 for Monteith et al.'s system's [12] skill from one judge (accompanied by a 5/10 from the other judge).

Taking the mean of the three qualities for each system, overall averages were 7.2 for Rahman & Manurung [17] and 6.8 each for Norton et al. [15] and Tearse et al. [22]. These observations indicate that Rahman & Manurung's system [17] was found to be more creative than the other systems, as it had a higher mean overall and the highest ratings for two out of three qualities. Its performance for appreciation, though, was only average (5/10). It could be argued that *DARCI* demonstrated a better all-round performance and was therefore found to be more creative.[10] The other two systems [4, 12] were considered less creative overall than these three systems, because they did not demonstrate clear abilities on some of the tripod qualities (given the information in the presentations on the systems). Of these two, Monteith et al.'s system [12] may have been slightly superior to that of Cook & Colton [4] because it demonstrated some aspects of both skill and imagination and received one rating of 7/10 (from Judge 2 for skill) in comparison to the rest of the ratings for these two systems (either left blank or rated as 5/10).

We can conclude that with the above use of the Creative Tripod, Rahman and Manurung's system performed best in creativity evaluation, followed jointly by Norton et al. [15] and Tearse et al. [22], then Monteith et al.'s system [12], and finally Cook & Colton's system [4].

[10] Given that Colton [2] does not investigate how to use the creative components for quantitative comparison, and that no such usage of the creative tripod was found to use as an example, exact conclusions are speculative. it is noted here that Colton's Creative Tripod is intended to identify whether computational systems can be considered candidates for potentially creative systems, rather than evaluating their creativity per se, so this case study stretches the application of the Creative Tripod somewhat beyond what Colton originally intended.

2.4 Collecting people's opinions of creativity

The evaluation results and feedback obtained for the 2011 case study were compared to human evaluations of the creativity of the case studies. The two judges were asked to say how creative each system was and their reasons and comments. Judges could choose from the following options to describe a system's creativity: Completely creative; Very creative; Quite creative; A little creative but not very, Not at all creative. More relevant to this paper's investigations, judges were also asked to rank the five systems in terms of relative creativity.

- The collage generator: 'Not at all creative/Quite creative'. The complexity of the processes used was praised by one judge but seen as trivial for creativity by the other.
- The poetry generator: 'A little creative but not very/A little creative but not very'. It generated interesting poetry but did not generate what was intended and was more aimed at generating a target example.
- *DARCI*: 'A little creative but not very/Quite creative'. *DARCI*'s ability to learn was highlighted as a useful attribute by one judge. The system may be more useful for interactive creativity with humans than as a standalone system, though, with one judge seeing the creativity of DARCI as located within the knowledge obtained from people.
- The story generator: 'A little creative but not very/Quite creative'. Whilst creating stories that seem to be fairly interesting but slightly nonsensical, the process did not seem to be optimised for increasing the interestingness of its stories, but for replicating as closely as possible a previous system (MINSTREL).
- The soundtrack generator: 'Quite creative/Quite creative'. Judges liked the intentions behind the system and its ability to combine two existing systems and layer human involvement, but needed more information for a fuller opinion.

An overall ranking of systems' creativity can be generated from the data on the judges' rankings, in Table 3. For each judge's rankings, 5 points were allocated to the system ranked most creative, down to 1 point for being ranked least creative. The two sets of ranking points were then summed together:

- Monteith et al. [12] was ranked most creative overall with 10 points (5+5).
- Norton et al. [15] was ranked second most creative overall with 7 points (3+4).
- Tearse et al. [22] was ranked third most creative overall with 5 points (3+2).
- Rahman & Manurung [17] and Cook & Colton [4] were ranked joint least creative overall with 4 points (3+1 and 1+3 respectively).

The ratings and feedback show some common opinions between the judges. For example, both judges praised the processes involved in Monteith et al. [12] and both criticised Rahman & Manurung [17] for their focus on replicating a target poem rather than creating new poetry. Individually, judges' opinions varied a great deal, as is perhaps to be expected with using only two judges who have differing backgrounds and expertise in the various creative domains covered.

One thing this has illustrated is that individual people can form very different first impressions of systems. Taking two people's opinions was useful for directed, constructive criticism, but was less useful for any significant statements about which systems are more or less creative than each other, compared to the more formal evaluation methodologies employed. Though a similar previous case study with 111 recipients [8] has suggested a larger number of judges does not necessarily produce conclusive distinctions between systems' perceived creativity, this evaluation for the 2011 case study showed the limits of what can be taken from a small number of judges.

Table 3. Ordering the case study systems by creativity: Judges' opinions.

Creativity	Judge 1	Judge 2
Most: 1	Monteith et al. [12]	Monteith et al. [12]
2	{ Rahman & Manurung [17] / { Norton et al. [15] / { Tearse et al. [22] (equal)	Norton et al. [15]
3	”	Cook & Colton [4]
4	”	Tearse et al. [22]
Least: 5	Cook & Colton [4]	Rahman & Manurung [17]

3 Comparing the success of different evaluation methods in the two case studies

Four different evaluation methods have been now used to evaluate the creativity of the five systems in the 2011 case study: SPECS (using the components in [7] as a basic definition of creativity); people's opinions of creativity of a system; Ritchie's empirical criteria and Colton's creative tripod. Each generated evaluative information on each system and comparisons of creativity between systems within each case study.[11] Here we focus on the relative rankings the evaluation methods generated for the five evaluated systems

3.1 Comparing evaluation results and feedback in the 2011 case study

No two methodologies produced the same results, but typically, *DARCI* [15] was judged one of the more creative systems and the collage generation module for *The Painting Fool* was judged one of the less creative systems. Otherwise, there was some disagreement between findings, largely caused by the lack of a 'ground truth' or baseline to judge the systems and the different domains that the systems work in. The priority in this case study in 2011 was on obtaining formative feedback for the system authors.

Here, we look at the results of each evaluation method to see if they can help us predict which systems have had a longer term contribution to the field of computational creativity. This is somewhat akin to the way in which we judge systems based on their presentation at a conference. While there are many factors beyond a conference presentation that determine what weight we give to a work's contribution, a major resource for information in computational creativity research comes from the information available at the key international conference for this research, the ICCC conference series. Both the papers and the presentations form key sources of information for computational creativity researchers; this is also true for the information in the 2011 case study.

Two ways in which we can measure contribution of papers to computational creativity research are (1) to count a paper's citations; and (2) to examine citations to see if those systems in 2011 have had direct influence in further computational creativity research.

[11] A full presentation and discussion of this information can be found in [8].

Table 4. Overall comparisons of the relative creativity of each system in the case study from 1: Most creative to 5: Least Creative. NB 'gen': 'generator'.

Evaluation Method	SPECS using components	Survey of opinions	Ritchie's criteria	Colton's tripod
1	*DARCI*	soundtrack gen	*DARCI*	poetry gen
2	story gen	*DARCI*	poetry gen	*DARCI* / story gen
3	poetry gen / soundtrack gen	story gen	collage gen	*DARCI* / story gen
4	poetry gen / soundtrack gen	poetry gen / collage gen	- (other two systems unrated)	soundtrack gen
5	collage gen	poetry gen / collage gen	- (other two systems unrated)	collage gen

3.2 Citation counts

Using Google Scholar, we can see how many citations the papers reporting each system under investigation have attracted since 2011 (reported in descending order of total citation count): see Table 5.

Table 5. Number of citations for each paper in the 2011 case study (in descending order of total citations, according to Google Scholar):

Paper	# citations	# non-self-citations
Norton et al. [15]	13	5 (38%)
Cook & Colton [4]	11	3 (27%)
Rahman & Manurung [17]	8	8 (100%)
Tearse et al. [22]	7	3 (43%)
Monteith et al. [12]	5	4 (80%)

Table 5 shows that Norton et al.'s work [15] has received the most citations overall, followed by Cook & Colton's work [4]. Monteith et al.'s paper [12] has to date received the fewest citations. If we look at non-self-citations, i.e. those citations from papers with no shared authors to the original paper, then Rahman & Manurung's work emerges as highest in influence both in terms of number of non-self-citations and overall percentage of non-self-citations compared total citations. At the other end of the scale, Cook & Colton's paper and Tearse et al.'s paper both attract only 3 non-self-citations.

We should remember the number of factors involved in citations: such as is the citation positive or negative? does the citation focus on the work cited or merely acknowledge it in passing? how active are the original authors in publishing their own work at other venues? But nonetheless, citation counts continue to be a key metric in measuring research impact. The absolute number of citations highlighted DARCI [15] (in rough alignment with the evaluation methods) and Cook & Colton's Painting Fool module [4] (not in alignment with the evaluation methods. Perhaps more importantly for this metric, the number of non-self citations highlighted Rahman & Manurung's poetry generator (in rough alignment with all the formal evaluation methods, though not the opinion-based evaluation), with Norton et al.'s paper on DARCI receiving the second highest number of non-self-citations (roughly aligning with all results from the 2011 study).

3.3 Direct influence on subsequent computational creativity research

What current (or successive) work did the 2011 papers inform? This is where citation data from both self-citations and non-self-citations can be investigated more thoroughly. We find that (in rough order of the 2011 case study rankings, across all evaluation methods):

- DARCI [15] is still being featured in subsequent publications in most years, with an active online community crowdsourcing data for DARCI's development. [15] is also cited as influencing work on other systems [14, 21, for example].
- Rahman & Manurung's poetry generator [17] has been cited across papers reporting multiple different pieces of work [13, 3, 21, for example]. The first author of this work has not since published work in the computational creativity field, unlike the other authors, but the work has clearly made some impact on the computational creativity field. The second author has since published work in computational creativity, but interestingly, has not since cited this 2011 paper.
- Tearse et al.'s story generator [22] has mostly been cited in papers considering further development of the MINSTREL reconstruction: showing influence in creativity development through one system, but not a wider impact (to date).
- Monteith et al.'s soundtrack generator [12] has been cited in reports of other systems, with some influence evident in the way the system is reported in some of these citations. The paper has not, however, been cited by the authors themselves, suggesting that development of this particular system has taken different paths since 2011.[12]
- Cook & Colton's collage generator [4] is arguably part of one of the most prominent systems longer term, being a module for the Painting Fool artistic system. This system has attracted much publicity through exhibitions, further publications, and public engagement/impact activities, though it is unclear whether the collage generator module is influencing this system, or whether it is a module that may or may not be used depending on the current application of the Painting Fool.

What we see here is that the DARCI system has again been recognised as valuable computational creativity software. Rahman and Manurung's poetry generator has also been found to have longer-lasting influence across computational creativity work, even though the lead author of this paper is not a regular participant in computational creativity research events.

The 'surprise' result given the 2011 case study results (when considered in isolation) is the long-lasting impact of Cook & Colton's collage generator. This reminds us that it is not merely an evaluation

[12] Perhaps somewhat ironically, two of the authors do however cite Rahman and Manurung's work in a later paper of theirs.

of a system which can help us judge the longer-term impact of a creative system in computational creativity research; there are several other factors in play such as the provenance of the system (e.g. its authors, what system(s) it is derived from). It is interesting, however, to see some consistency in alignments between the 2011 evaluations and the metrics employed here for longer term impact.

4 Summary

The 2011 case study carried out during the development of the Standard Procedure for Evaluating Creative Systems (SPECS), showed how various computational creativity evaluation methods could be applied to evaluate the creativity of various types of creative systems from different creative domains. This 2011 case study captured first impressions and initial evaluations of how creative systems were, with limited information and resources, and under time pressures.

Analysis of these evaluations provided information about how creative the systems were perceived to be and what information contributed to this, relative to the creative domain. This 2011 case study also highlighted what information is most useful to help people make evaluations of creativity based on conference papers and presentations - key sources of information for computational creativity researchers. Several evaluation methods were applied to the systems evaluated in the 2011 case study. As well as SPECS [7, 8], people's opinions were consulted on the creativity of the systems. Two key existing methodologies for computational creativity were also applied: [18, 2, Ritchie's criteria and Colton's creative tripod, respectively]. Results were compared; it was noted that few 'right answers' or 'ground truths' for creativity were found in the 2011 case study.

The consequences of judging a system given limited and perhaps incomplete information meant that occasionally important information for evaluation is missing. This affected the use of all the evaluation strategies employed. It is interesting to see which methodologies were most robust when dealing with missing information. Collecting people's opinions seemed the best approach at coping with missing information, as might be expected given that little was specified for the humans to use as a definition of creativity. SPECS was relatively robust, as was Colton's tripod framework. Ritchie's criteria approach was the most affected by missing information, as various criteria could not be applied and the absence of information on inspiring sets and example outputs had significant effects.

Looking longer term at whether initial evaluations of creativity can help us predict which systems are likely to make a longer term contribution to creativity: this has been an interesting experiment. Certainly, some alignment was found between the 2011 evaluation results and the impact five years on that each system/paper has made, as can be measured by citation quantity and types. However the evaluation results did not directly correlate with study of later impact. The creativity of systems presented in computational creativity is one factor which contributes to their value to the community, but as discussed above, it is of course not the only factor. However, evaluation methods are giving us some hints for gauging longer term impact. This experiment has only looked at impact over a five year period. Perhaps in ten years (at AISB'21?), or over even longer time periods, we will uncover different results?

REFERENCES

[1] Anna Bentkowska-Kafel, 'The Fix vs. the Flux: Which digital heritage?', in *Netpioneers 1.0 - archiving, representing and contextualising early netbased art*, eds., Dieter Daniels and Günther Reisinger, 145–162, Sternberg Press in association with the Ludwig Boltzmann Institute, Berlin, Germany / New York, NY, (2009).

[2] Simon Colton, 'Creativity versus the Perception of Creativity in Computational Systems', in *Proceedings of AAAI Symposium on Creative Systems*, pp. 14–20, (2008).

[3] Simon Colton, Jacob Goodwin, and Tony Veale, 'Full-FACE poetry generation', in *Proceedings of the International Conference on Computational Creativity*, pp. 95–102, Dublin, Ireland, (2012).

[4] Michael Cook and Simon Colton, 'Automated Collage Generation - With More Intent', in *Proceedings of the 2nd International Conference on Computational Creativity*, pp. 1–3, Mexico City, Mexico, (2011).

[5] Chris Darwin. Obituary: Christopher Longuet-Higgins. *The Guardian*, June 10th, 2004, jun 2004.

[6] Neil Chue Hong, Steve Crouch, Simon Hettrick, Tim Parkinson, and Matt Shreeve, 'Software Preservation: Benefits framework', Technical report, Software Sustainability Institute and Curtis & Cartwright Consulting Ltd., (dec 2010).

[7] Anna Jordanous, 'A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative', *Cognitive Computation*, **4**(3), 246–279, (2012).

[8] Anna Jordanous, *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*, Ph.D. dissertation, University of Sussex, Brighton, UK, sep 2012.

[9] Hisar Maruli Manurung, *An Evolutionary Algorithm Approach to Poetry Generation*, Ph.D. dissertation, School of Informatics, University of Edinburgh, Edinburgh, UK, 2003.

[10] James Meehan, 'Tale-Spin', in *Inside computer understanding: five programs plus minatures*, eds., R C Schank and C K Riesbeck, Lawrence Erlbaum Associates, Hillside, NJ, (1981).

[11] James Richard Meehan, *The metanovel: writing stories by computer.*, Ph.D. dissertation, Yale University, New Haven, CT, USA, 1976.

[12] Kristine Monteith, Virginia Francisco, Tony Martinez, Pablo Gervás, and Dan Ventura, 'Automatic Generation of Emotionally-Targeted Soundtracks', in *Proceedings of the 2nd International Conference on Computational Creativity*, pp. 60–62, Mexico City, Mexico, (2011).

[13] Kristine Monteith and Tony Martinez, 'Automatic generation of melodic accompaniments for lyrics', in *Proceedings of the 3rd International Conference on Computational Creativity*, Dublin, Ireland, (2012).

[14] R G Morris, S H Burton, P M Bodily, and D Ventura, 'Soup Over Bean of Pure Joy: Culinary Ruminations of an Artificial Chef', in *International Conference on Computational Creativity*, p. 119, (2012).

[15] David Norton, Derrall Heath, and Dan Ventura, 'Autonomously Creating Quality Images', in *Proceedings of the 2nd International Conference on Computational Creativity*, pp. 10–15, Mexico City, Mexico, (2011).

[16] Jonathan A Plucker and Ronald A Beghetto, 'Why Creativity is Domain General, Why it Looks Domain Specific, and why the Distinction Doesn't Matter', in *Creativity: From Potential to Realization*, eds., Robert J Sternberg, Elena L Grigorenko, and Jerome L Singer, chapter 9, 153–167, American Psychological Association, Washington, DC, (2004).

[17] Fahrurrozi Rahman and Ruli Manurung, 'Multiobjective Optimization for Meaningful Metrical Poetry', in *Proceedings of the 2nd International Conference on Computational Creativity*, pp. 4–9, Mexico City, Mexico, (2011).

[18] Graeme Ritchie, 'Some Empirical Criteria for Attributing Creativity to a Computer Program', *Minds and Machines*, **17**, 67–99, (2007).

[19] David Robey. Introduction to Digital Humanities. Talk, Sep 2011.

[20] Jeff Rothenburg. Ensuring the Longevity of Digital Information. Available at http:// www.clir.org / pubs / archives / ensuring.pdf (last accessed November 2012), feb 1999.

[21] M. R. Smith, R. S. Hintze, and Dan Ventura, 'Nehovah: A neologism creator nomen ipsum', in *Proceedings of the 5th International Conference on Computational Creativity*, Ljubljana, Slovenia, (2014).

[22] Brandon Tearse, Peter Mawhorter, Michael Mateas, and Noah Wardrip-Fonin, 'Experimental Results from a Rational Reconstruction of MINSTREL', in *Proceedings of the 2nd International Conference on Computational Creativity*, pp. 54–59, Mexico City, Mexico, (2011).

[23] Scott R Turner, *The creative process: a computer model of storytelling and creativity*, Erlbaum, Hillside, NJ, 1994.

A Computational Framework for Aesthetical Navigation in Musical Search Space

Sahar Arshi[1], Darryl N. Davis[1]

Abstract. This article addresses aspects of an ongoing project in the generation of artificial Persian (-like) music. Liquid Persian Music software (LPM) is a cellular automata based audio generator. In this paper LPM is discussed from the view point of future potentials of algorithmic composition and creativity. Liquid Persian Music is a creative tool, enabling exploration of emergent audio through new dimensions of music composition. Various configurations of the system produce different voices which resemble musical motives in many respects. Aesthetical measurements are determined by Zipf's law in an evolutionary environment. Arranging these voices together for producing a musical corpus can be considered as a search problem in the LPM outputs space of musical possibilities. On this account, the issues toward defining the search space for LPM is studied throughout this paper.

1 INTRODUCTION

The *Radif* is the repertory of Persian traditional music which consists of different *Dastgāhs* [1]. Dastgāhs are distinguished from each other by their musical modal systems and the movement of melodies[2], [3], [4]. Dastgāhs have been unevenly mapped to modes in Western musical terminology [1]. The Dastgāh concept determines both the title for a group of individual pieces with their characteristic modal identity and the primary mode in each group [1]. There are twelve principle groups of modes in Persian music, namely, Shur, Abou'atā, Bayāt-e-Tork, Afshāri, Dashti, Homāyoun, Bayāt-e-Esfehān, Segāh, Chāhārgāh, Māhour, Rāstpanjgāh, and Navā [1]. Each Dastgāh consists of individual melodies called Gushé, which vary in length and importance[1]. Performing in a Dastgāh begins with Darāmad which are the most representative pieces of a Dastgāh. Darāmads have the prominent mode and melodic patterns of the Dastgāh itself giving the Dastgāh its identity [1], [5]. The modulation occurs with the move from one Gushé to another or a change in the central tone, or Shāhed note [6].

The current musical warehouse is the result of centuries of evolution of Persian music conjoint with historical and cultural transmutations. However, there are still varieties of other melodies waiting to emerge. Once modulated with Western music it can be considered as a potential bed for cross cultural interactions. Although Persian music has vast musical systems in comparison to its Western contemporary music counterpart, one of the problems encountered is the entrapment in the structures. This makes the composition more reliant on the emergence of great masters whom with their novel creativity and familiarity of the complexities of Persian music were able to put a step forward in this field and add pieces to different Dastgāh. Therefore the variety of melodies and Gushé in a Dastgāh is limited to what was produced in the past.

The use of algorithmic composition has been under investigation for many years with different motivations: Mechanization of music production; exploration of the behaviour of the algorithms; mathematical models in generating the patterns; studying the cognitive behaviour of creation in human being [7]; and modelling biological patterns in nature in respect to music.

Mechanisation of music generation has been done for producing melody, rhythm, harmonization, and counterpoint or imitating a specific genre of music or composition style [8]. The level of automation varies from generating motifs for inspiration to more complex corpus composition. Computer aided algorithmic composition is the term applied for assisting musicians in the composition process and providing them with new materials; (some available frameworks or languages for making musical software include Csound [9], MAX/MSP [10], while some musical software include EMI [11], [12], GenJam [13], and LBM [14], [15]). Deeper levels of composition automation target minimal or no interactions with human (Melomics corpus generation [16])

Methodologies in algorithmic composition can be categorized, based on the survey from [8], in four groups: , knowledge based systems; machine learning; evolutionary algorithms; and computational intelligence (e.g. cellular automata). All of the aforementioned categories except the last one apply human knowledge in their application. They have been widely used both for style imitation and creating novel music. However, cellular automata are able to generate novel material without utilising existing human domain knowledge. This potential of creativity makes them well-suited for exploring new dimensions of music composition.

There have been good progress with the research into genre imitation; successful applications include Strasheela [17]. Most research efforts are now focused on algorithmic creativity applications. The future directions for algorithmic composition includes hybrid methods [8] that use cellular automata (CA) as their music generator.

Liquid Persian Music is a CA based toolkit for exploring various musical possibilities. Pattern matching rules classify output from the cellular automata and update the parameters of a synthesiser to yield audio output. Controlling synthesizer parameters by means of the emergent nature of CA is an important characteristic of LPM. In this work each parameterisation of LPM, across both CA and pattern matching rules and the synthesiser, is considered to be an audio voice. Sequencing LPM produced

[1] Dept. of Computer Science, Univ. of Hull, HU6 7RX, UK.
Email: { S.Arshi@2014., D.N.Davis@ }hull.ac.uk

voices in a musical manner requires investigating a huge search space. The dimensions of this space are defined by the number of CA rules, pattern matching rules, the elements of sound synthesizer and melodic structure. One important question that this research addresses is how to evaluate the musical productions of such system in terms of aesthetics? Furthermore, is there a measurement for the creativity of the system itself?

Creativity is a diverse concept with multiple definitions. We need to be specific with its definition in order to evaluate it in the current project. In the next section, more clarifications addressing the concept of creativity are presented.

Previous researches show the application of musical Turing tests [18] and surveys [19] for evaluating musical productions; however, giving equal measurements for creativity for human and machine is challenging [8]. These evaluations are done according to pre-existing musical knowledge and cultural backgrounds, while the real creativity goes beyond pre-existing musical styles [20]. Despite the widespread research in the area of algorithmic composition itself, less attention has been given to assessing the outputs from creativity viewpoints [20]. Evaluating the computational creativity can be traced in [20] , [21], [22]. Nevertheless, the creativity of an artefact can be perceived by their aesthetical values [20]. Various scientific studies have been conducted on the matter of universals for recognizing natural or human-like phenomena, as well as frequency distributions, and power law. Among these is the use of Zipf's law as a basis for aesthetical measurement [7]. Zipf's law has had successful applications in measuring the aesthetical aspects of music [23] and we have been looking at it as the start point for advancing the current research from an aesthetical point of view.

In a previous experiment [24] the LPM output voices were analyzed in the search for finding proper tools for enhancing them in a musical way. A pool of voices have been produced and the pleasantness of each of those elements have been evaluated against aesthetic measurements using Zipf's law [23]. In a later experiment in the same paper, random sequences of voices were produced with nearly acceptable Zipf's slopes. The next level of investigations consists of designing a computational framework for sequencing LPM outputs in an evolutionary environment. The idea is that Genetic Algorithms are suitable candidates for delving in the problem of sequencing LPM musical elements. However, the huge search space makes it an impractical one, unless suitable constraints are taken into account.

In the second section, different types of creativity from viewpoint of computation are described. The third, fourth, and fifth sections revolve around background studies relevant to the current research. These include cellular automata and optimization methods and their applications in algorithmic composition and Zipf's law. In the sixth section the features of LPM software in employing cellular automata is briefly overviewed and the basis for measuring the creativity of the sequencer in the evolutionary domain is presented. The seventh part is devoted to the design of computational framework that allow the sequencing of LPM productions to be viewed as a search problem. The issues raised and potential solutions are discussed in detail. The paper is concluded with future direction of the research.

2 ON COMPUTATIONAL CREATIVITY

"What is creativity?" –This can be considered as an open-ended philosophical question. There are no boundaries for creativity, yet binding creativity in a framework for a definition is a necessary but difficult task. However, an artefact has some representative features which describe its qualities to some extent. These qualitative descriptions clarify the attributes an artefact should have to be considered as a piece of art work. Amongst all descriptions what is clear is that art and novelty have been two inseparable concepts. Sometimes a black dot on a white canvas is defined as a masterpiece and is exhibited in art galleries. The work of John Cage in his composition "four minutes and 33 seconds of silence" unbounds framed viewpoints towards art and creativity with avant-garde music. In a silent musical performance he lets the energy from audience noise vibrate the strings of a grand piano. The interaction of audience noises and musical instruments is popular as aesthetics of art performance. There are other criteria for defining creativity other than novelty, for example quality [21]. This discusses how the creation is to be considered to be a high-quality instance of its genre. Jon McCormack defines this attribute of creativity as being exhibitable [7].

Two different viewpoints exist about man-machine creativity. The machines that create art-like productions, and the machines which are autonomous in creating art [7]. The aim of creating could be to satisfy an audience or could involve the exploration of general meaning of creativity, without contributing to human comprehension or appreciation.

Boden [22] defines three types of creativity: combinational; exploratory; and transformational. She states all can be modelled by artificial intelligence. Combinational creativity consists of populating pre-existing materials and linking them in an artistic manner for generating new ideas. Exploratory creativity includes navigating in a *conceptual space* with implicit constraining rules. This exploration can result in discovering new transformed styles which would not have existed before an alteration happening on one or more of their defining dimensions (transformational creativity).

3 THE NATURE OF CELLULAR AUTOMATA & ITS APPLICATION IN ALGORITHMIC COMPOSITION

The advent of cellular automata originally dates back to 1940s, when Von Neumann was looking forward to develop a system capable of reproduction, comparable in certain respects with biological breeding [25], [26]. Cellular automata was studied as a dynamical system in 1960s [27]. Cellular Automata are discrete dynamical systems whose global intricate behaviour is determined by the reciprocal influence of identical elementary individuals.

Cellular automata exhibiting myriad genres of behaviour have been targeted as a creative tool for artists. By increasing the number of states and neighbourhood size, the state space expands exponentially, in a way that the normal life expectancy of a human is not adequate for navigating through all these patterns. Amongst all the various applications such emergent machine can have are, namely, the extraction of overall conformation for composition, MIDI sequencing, and sound synthesis [28].

Two of the early models of musical cellular automata include Beyls cellular automata explorer, and CAM developed by Millen. Having the aim of achieving complex musical patterns in the output [8], Beyls investigated broad criteria of configurations for

CA rules, and cell neighbourhood [29]. Some of these include the application of time dependent rules, and involving the neighbour states from previous and future generations in the computations. Dale Millen employed two and three dimensional game of life cellular automata and mapped the results to pitch and duration. He later explored the formation of musical organization from CAM [29].

Other popular cellular automata musical systems are CAMUS and Chaosynth [30], [31]. CAMUS exploits Game of Life and Demon Cyclic Space, and uses a Cartesian space mapping to MIDI for achieving musical triplets. The main idea in CAMUS is to model the dissemination of musical patterns in time by simulating the same effect in cellular automata [31]. Chaosynth is based on the model of chemical reactions of a catalyst. It is a cellular automata sound generator based on the production of sound granules which are the results of underlying additive synthesis processes. However the produced tones do not often resemble the acoustic sounds found in the real world; they are sometimes reminiscent of the natural sounds flow as well as the sound of waterfalls, or insects [30]. The interested reader is referenced to [8], [29] for a thorough review on previous research on the application of CA in generating electronic music.

Cellular Automata are usually used as a hybrid tool beside other artificial intelligence tools in music composition algorithms, since, in isolation, they do not presently produce melodic sounds. However, they can be a source of raw material and structures for inspiration for musicians [8]. In the end the generated sounds may need heavy editing by the composer and so be conformed to musical playing as stated by Xenakis; one of the pioneers who used CA for achieving the general structure of his compositions [32], [19]. Similar issue have been stated by Miranda, the creator of CAMUS, who considers the results as not being very musical [33].

4 GENETIC ALGORITHMS & THEIR APPLICATION IN ALGORITHMIC COMPOSITION

Genetic Algorithms (GA) are a class of Evolutionary Algorithms inspired by natural selection [34]. They are employed in areas of search and optimization. Previous applications of GAs imply their success in problem solving for domains with widespread solution spaces [35]. Therefore, they can be considered a well suited candidate in music composition, with its almost infinite possible combinations of musical elements. However, in order to guide the search and constrain the musical search space one can tailor fitness functions which fulfil musical aesthetical aspects or adhere to certain musical tastes [36].

A population of individuals are randomly initialized in a mating pool. Candidate solutions are coded as genotypes and are continually evolved in each nascent generation. The solutions contribute to crossover and mutation operations according to their fitness function. This assessment guarantees the survival of the most competent genes and raises the expectancy of convergence to optimal solutions. The reproduction operation consists of the selection of parents as the fittest individuals for breeding, which then undergo the crossover and mutation operations. In crossover, individual parents are selected and their genes are transmitted to each other by swapping, mostly in a meaningful manner. The mutation is a low-probability operation and involves changing a gene in the genotype [37]. It can help the search by avoiding being entrapped in local solution spaces. The algorithm stops when a pre-specified goal has been satisfied or some sort of limitations such as time or number of generations has been reached [36].

In previous applications evolutionary algorithms have been widely used for composing melodies, and harmonizing pre-specified melodies. The fitness function can be interactive or autonomous. In interactive fitness functions a human user evaluates the candidate individuals in the population. These fitness functions usually contribute to user fatigue and should be used in domains where other fitness functions are unable to gain the desired results. The other types of fitness assessment usually contribute to the application of machine learning methods. In the following some examples of both types of fitness functions are described.

Horner and Goldberg [38] are one of the first to present the application of genetic algorithms in algorithmic composition. Thematic bridging is a composition method; starting from an initial pattern, the system goes into a series of evolutionary process to transform to the final pattern. The GA individuals are the transformation operators and the fitness function is evaluated as the distance to the target pattern. The sequences of the generated patterns are the output of the system. Jacob [39] applied three phase modules in the design of his composition system; the *Ear*, *Composer* and the *Arranger*. The human user trains the *Ear* which acts as an evaluator in the process of creating musical motifs according to authorized intervallic combinations. The *Arranger* is determined by the user as well, to reorder and assemble the output in the form of musical phrases. In GenJam [40], Biles devised an evolutionary algorithm for generating Jazz melodies. Later, he used an artificial neural network (ANN) to automate the task of evaluation to overcome the interactive fitness function bottleneck. However, the ANN failed to extend the evaluations to cases other than what was in its training set [8].

Genetic algorithms have been applied independently or as hybrid models accompanying various self–governing artificial intelligence and computational methods as well as knowledge-based models, Markov chains, artificial neural networks, and complex systems in producing artificial music.

Fitness functions can be defined simply as a weighted sum of distances to a target melody, however, if the musical statistical are selected poorly, reaching satisfactory results are unlikely to happen [41], [42].

In a series of applications neural networks have been used as fitness functions. Neurogen applies two neural networks, one for assessing the intervals between pitches, the other one for evaluating the overall structure. One of the successful neural networks and genetic algorithms hybrid approaches in computer music is the work of Manaris and his colleagues. Manaris et al. trained neural networks as a fitness function with statistical metrics to identify individual compositions with Zipf's distribution property [23].

Markov Chains have been applied as fitness functions for evolving musical sequences in a number of applications [43]. In [44] variations between two musical pieces have been modelled using random jumps between two Markov chains trained with two different pieces. Hidden Markov models trained with proper counterpoint training set have been able to produce Palestrina style first species counterpoint [45]. HMM trained with chorale harmonization add extra voice elements to a pre-processed melody in [46].

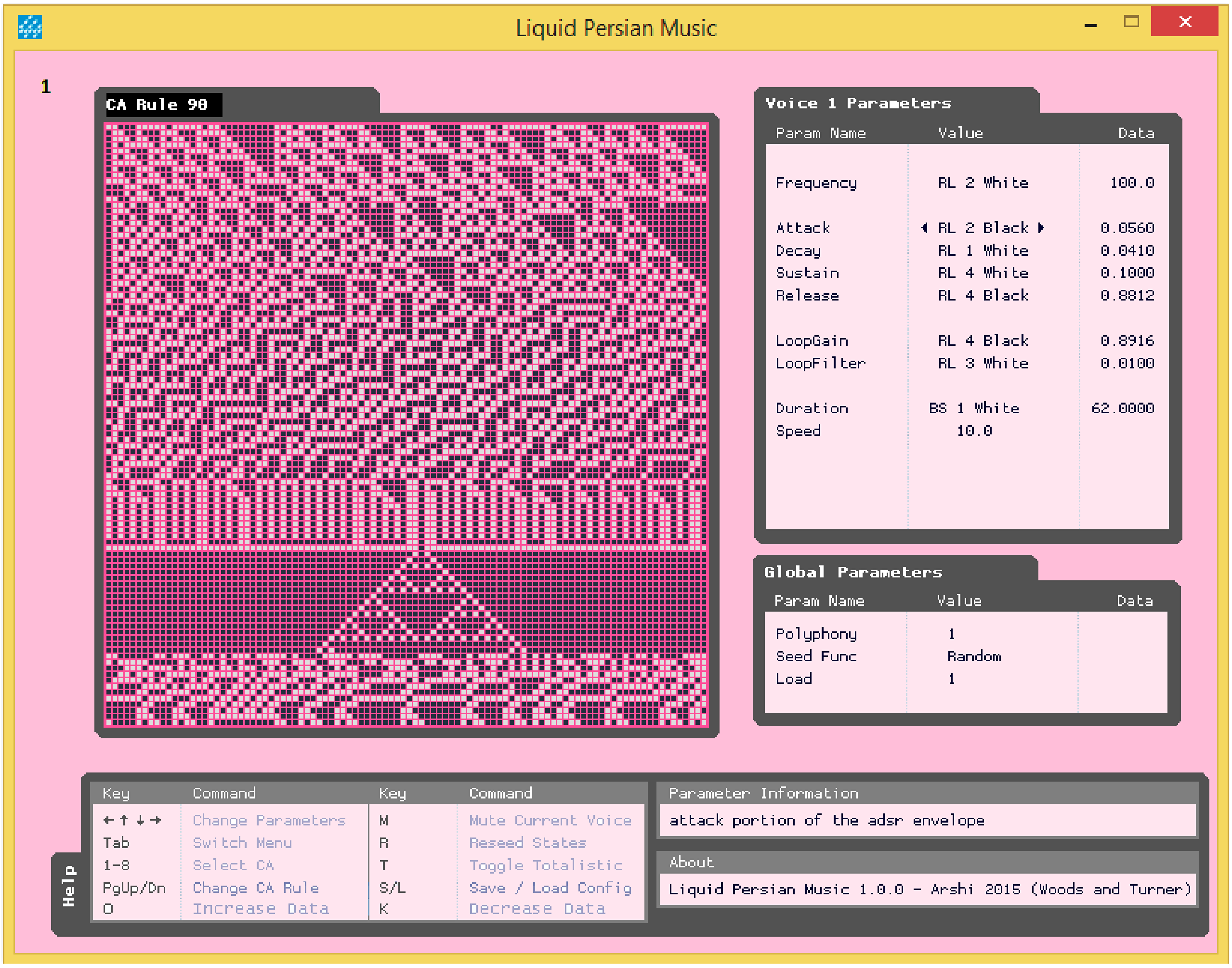


Figure 1. LPM user interface.

In [19], n-gram models, Zipf's law, and information entropy are applied as trainable fitness functions in a series of experiments. Musical samples are used to train N-gram classifier which is later applied as the fitness function in a random mutation hill climber. These fitness functions evaluate sequences of pitches, and the genetic operators are employed as tools for search space navigation. Later in the same work evolutionary algorithms are applied to evolve cellular automata as a music generator.

5 ZIPF'S LAW IN MUSIC

Zipf's law characterizes the scaling attributes of many natural effects including physics, social sciences, and language processing. Events in a dataset are ranked (descending order) according to their prevalence or importance [23]. The rank and frequency of occurrence of the elements are mapped to a logarithmic scale, where linear regression is applied to the events graph. The slope and R^2 measurements demonstrate to what extent the elements conform to Zipf's law. A linear regression slope of -1 indicates Zipf's ideal. Zipf's law can be formulated as $F{\sim}r^{-a}$, in which r is the statistical rank of the phenomena, F is the frequency of occurrence of the event, and a is close to one in an ideal Zipfian distribution. The frequency of occurrence of an event is inversely proportional to its rank. $P(f) = 1/f^n$ is another way to express the Zipf's law. $P(f)$ is the probability of occurrence of an event with rank f. In case of n=1(Zipf's ideal), the phenomenon is known as pink noise. The cases of n=0 and n=2 are called white and brown noises, respectively.

Voss and Clarke [47] have observed that the spectral density of audio is 1/f like and is inversely proportional to its frequency. They devised an algorithm which used white, pink, and brown noise sources for composing music. The results show that pink noise is more musically pleasing due to its self-similarity characteristics, the white noises are too random, and the brown noises are too correlated producing a monotonous sound.

In the musical domain, Zipf's metrics are obtained by enumerating the different musical events' frequency of occurrence and plotting them in a log-log scale versus their rankings. The slope of Zipf's distribution ranges from $-\infty$ to 0. The decreasing of the slope to minus infinity reflects an increase in the level of monotonicity. The r-squared value is between 0 and 1. Various publications explore the utilization of Zipf's law in musical data analysis and composition. Previous experiments show its successful application in capturing significant essence from musical contents. In [23] the Zipf's metrics consist of simple and fractal metrics. The simple metrics include seventeen features of the music as well as the ranked frequency distributions of pitch, and chromatic tone. Fractal metrics gives a measurement of the self-similarity of the distribution. These metrics were later used to train neural networks to classify musical styles and composers, with an average success rate of over ninety percent; demonstrating that Zipf's metrics extract useful information from music in addition to determining the aesthetical characteristics of music pieces.

6 LPM OVERVIEW

Liquid Persian Music (LPM) is an auditory software tool developed at the University of Hull [15][48]. LPM explores the idea of artificial life systems in producing voices which can be considered as new types of electronic music. The software takes advantage of the Synthesis Toolkit (STK) [49] for implementing the physical model of a stringed musical instrument. A model of its parameters are controlled by defined pattern matching rules. An OpenAL library is responsible for propagating the producing voices. Figure 1 illustrates the LPM user interface.

The elementary CA used in LPM consists of an assembly of cells arranged in a one dimensional array that produces a two dimensional matrix over time. Each cell is in one of k finite states at time t, and all the cells evolve simultaneously. The state of a cell at time *t* depends on its state and its neighbours' states at time *t-1*. In the one dimensional elementary CA (which is the subject of this study), the permutations of each cell with its two adjacent neighbours specifies eight situations. Once allocated to binary states, the selection of one of the 256 local transition rules specify the CA evolution [27]. Wolfram studies on CA recognize four classes of behavior, namely, fixed, cyclic, chaotic, and complex. Li and Packard [50] subdivided the second class to three further subgroups, namely heterogeneous, periodic with intervals greater than one, and locally chaotic.

In every time step of the CA, the pattern matcher extracts the difference between consecutive generations. Twenty different pattern matching rules have been defined in this software as well as metrics using Dice's coefficient, and Jaccard similarity. The obtained values from the pattern matchers are then fed as parameters into the STK synthesizer for producing sounds. Some of the synthesizer parameters include ADSR envelope, loop gain, and the musical instrument string length for defining frequency. Further information about the software can be found in [51].

An important point is that the aggregation of a CA rule and a pattern matching rule on each of the synthesizer elements does not produce a single note but a collection of notes; these are referred to as *voices* throughout this paper.

Studying the musical behaviour derived from one-dimensional (1D) CA does not require the investigation of the 256 rules' behaviours. The rule space can be reduced to 88 fundamental behaviours [52] by applying conjugate, reflection, and both transformations together [27], since they lead to rule sets with inherently equivalent behaviour. (The interested reader is referred to [27] for formulation of conjugate and reflection transformations and how they are applied to find equivalent CA rules). The 88 1D CA rule behaviours, 7 defined synthesizer parameters, together with 20 pattern matching rules, expand the number of voices to $88 * 20^7$. If the pattern matchers are chosen from separate cellular automata rules, then the voices number would become $88^7 * 20^7$. Considering the temporal and intervallic patterns and the CA number of iterations the search space would expand to $88^9 * 20^9 * t$.This defines the base auditory search space for the computational framework being developed.

In [24] the outputs of LPM have been explored through graphs and auditory tests. The behaviour of each of the pattern matching rules over one-dimensional cellular automata rule space have been explored and categorized in an initial step. The consequent experiments in [24] focus on the study of Zipf's law on LPM individual voices and sequence of voices.

In a first experiment, the output distributions have been investigated regarding their compliance with Zipf's law. Figure 2 presents examples on LPM output with their corresponding Zipf's distribution. In a second test, the results from the first experiment were categorized according to the expectations from studying their behaviours. This part of the experiment was conducted by a confusion matrix to measure the convenience of using Zipf's law for recognizing musical from unmusical voices. In a third experiment, collections of voices were sequenced; some with pleasing Zipfian slopes results that were expected to have musical voices. Figure 3 depicts two samples of Zipfian distribution for random sequences of voices. The random sequences of voices are selected from pattern matching rules applied to CA with iterations up to 10000 and 20. Figure 3 (a) shows a more monotonous output than figure 3 (b). The sequence of longer motives seem to have a more tedious structure. However some of the CA and pattern rules did not contribute to musical outputs by themselves. However, experiments with crafted pieces have shown that the proper combination of the voices can produce acceptable musical results. The measured Zipfian slopes characterize the global features [23] of the produced music. The attention was kept on one dimension of synthesizer (the frequency) and on global measurement of aesthetics throughout the study, for simplicity. In the next section we shall reveal some of the challenges in designing a computational framework that will allow candidate LPM voices to be sequenced into musical composition system. The experiments conducted in the previous paper [24] have been targeted as a base for designing the fitness function for the problem of sequencing LPM voices in an evolutionary space.

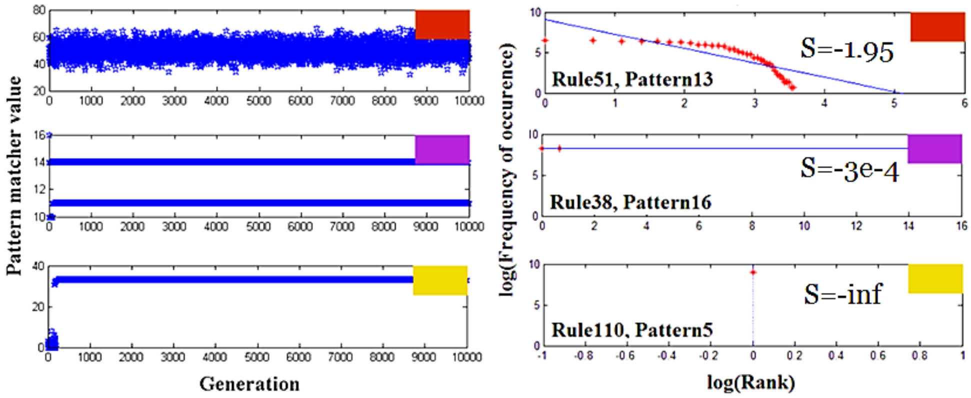
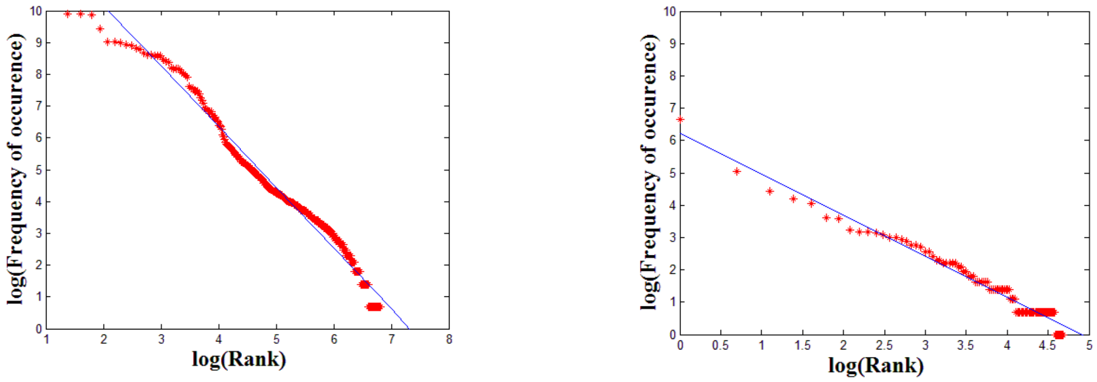


Figure 2. Some Examples for LPM Outputs and Zipf's distribution.



(a) S = -1.91 , R^2 = 0.96 (b) S= -1.36 , R^2 = 0.97

Figure 3. Zipfian distributions of random sequenced voices with lengths up to (a) 10000 and (b) 20.

7 THE DESIGN OF LPM SEQUENCER

In this section, sequencing LPM voices is taken as a search problem for producing the required melodic structure. Designing such a system gives raise to the following questions:

- How to design an efficient search space traversal which resolves the sequencing problem within the constraints of given hardware resources.
- How to sequence voices in a musical manner? What are the defining musical critiques?
- What are the possible genotypes and phenotypes of a musical sequencer based on LPM?
- Is there a measurement for the creativity of the system itself?

Applying Genetic Algorithms for search and optimization of musical sequences has special requirements. For example, defining the search space; specifying the musical knowledge and rule representation; and the choice of an appropriate fitness function [36]. The search for finding optimal solutions is guided by assigning higher fitness to competent individuals. Since there are, in effect, infinite possibilities for producing music; it is necessary to define suitable constraints for limiting the search space.

As stated in the previous section LPM outputs are a set of voices instead of notes. The voices resemble musical motives of varying lengths depending on the number of cellular automata iterations involved in their production. The design of competent genotypes and phenotypes are requirements for an efficient search. The genotypes are codes which manifest a higher level of behaviour known in the phenotype. For example the eye colour is coded in genes. However, what is seen as blue, green, and etc., are the phenotypes. It should be noted that in the LPM system, the phenotypes are the voices which are heard as the behaviours of the individuals and the genotypes are the set of genes coded whether as binary or integer representations.

A first naïve design for the search space would be to define the individuals as the elements of voices set. Regarding the huge search space and our current facilities, software implementation is nearly impossible unless the search space is reduced by a notable amount. Perhaps selecting a limited number of voices and evolving them would be a more feasible solution. During the evolution of such a design, all the contributing parameters change dynamically to a point that fulfils predefined musical expectations or at least tries to do so. This stabilization includes a gradual justification of musical parameters and general improvement in each generation. There are no unique solutions to musical problems, In fact starting from the same initial conditions, the search may result to different sets of solutions in every execution.

Further improvement in the design is to divide the search problem into several multi-optimization ones, relating to the constituent elements of the produced melody based on the LPM output. The first search determines the structure of the melody, including the pitch frequency, the intervals and the note durations. The second search problem involves the optimization of the remaining synthesizer elements. This separation provides two categories of different natures for exploration. The search pool sizes of which becomes $88^2 * 20^2 * t$ and $88^7 * 20^7$ respectively. Evaluators for the individuals of each of these search spaces vary. This paper focuses on the first optimization problem though. The related fitness function scores every one of the individuals based on their statistical aesthetical competence, coded in their individual genes.

For crossover and mutation after selection operator, various methodologies can be thought of. The crossover operator can be defined as swapping the codes of the related voice producer parameters. By this methodology it is guaranteed that the newly born individuals are those previously existing in the grand pools which are given the chance of being investigated musically towards the aims of the genetic algorithm.

8 CONCLUSIONS & FUTURE WORK

CA evolution have been employed as a controller for the parameters of a synthesizer. Computational intelligence models as well as cellular automata are sources of creativity which can produce musical material without contributing to human knowledge. This research requires working with exploratory, and transformational types of creativity. Evolutionary algorithms have been found to be well-suited for this kind of navigations. Genetic algorithms have been chosen as a creativity exploratory tool for evolving sequence of voices.

LPM software, equipped with cellular automata and synthesis tool kit, has been introduced as an assisting tool for producing music. This paper provides a conceptual approach towards the design of a computational framework for sequencing LPM voices. We have described the problem of sequencing voices from a creativity point of view. Some existing visions towards computational creativity have been discussed. The dimension of the search space have been determined regarding the number of elements involved in voice generation and the components related to producing the melody. The search space is then divided to different categories regarding their nature as two different optimization problems. These include the psychoacoustic and melodic structure of LPM output. We are developing an evolutionary environment to enable this. Aesthetical measurements based on Zipf's law have been propounded as a base for designing fitness function for the optimization problem. Although, Zipf's law can be considered as a good approach for investigating the pleasantness of the output melody, there are other approaches which can be taken into account. Experiments of this kind (measuring the frequency distribution) are to be necessary but not sufficient conditions for investigating the aesthetical aspects of the phenomena (music in our case). However, they have been taken as an integral part in the design of the fitness function in the first stage. The next level of evaluation could contribute to human auditory tests in the form of survey.

The future research direction includes the design of fitness function for the multi optimization problem of sequencing LPM outputs.

REFERENCES

[1] H. Farhat, *The Dastgah Concept in Persian Music*. Cambridge University Press, 1990.
[2] E. Zonis, "Contemporary Art Music in Persia," *Music. Q.*, vol. 51, no. 4, pp. 636–648, 1965.
[3] J. During, "Dastgah, Encyclopedia Iranica," *Iranica*. 2011.
[4] "Radif (music)," *Wikipedia, the Free Encyclopedia*. .
[5] L. Englander, C. Lecocq, and G. Kerker, "The Radif of Persian Music : Studies of Structure and Cultural Context by Bruno Nettl ; Carol M . Babiracki ; Béla Foltin; Daryoosh Shenassa ; Amnon Shiloah ; Accompanying Cassette : The Persian Radif in Comparative Perspective by Bruno Nettl ; La musiq," 2015.
[6] P. Heydarian and J. D. Reiss, "The Persian Music and the Santur Instrument," *Proc. 6th Int. Soc. Music Inf. Retr. Conf.*, pp. 524–527, 2005.
[7] J. McCormack, "Facing the Future: Evolutionary Possibilities for Human-Machine Creativity," *Art Artif. Evol. A Handb. Evol. Art Music*, pp. 417–451, 2007.
[8] J. D. Fernández and F. Vico, "AI methods in algorithmic composition: A comprehensive survey," *J. Artif. Intell. Res.*, vol. 48, pp. 513–582, 2013.
[9] Boulanger, *The Csound Book: Perspectives in Software Synthesis, sound design, signal processing, and programming*. MIT Press, 2000.
[10] M. Puckette, "Max at seventeen," *Comput. Music J.*, vol. 26, no. 4, pp. 31–43, 2002.
[11] D. Cope, "An expert system for computer assisted music composition," *Comput. Music J.*, vol. 11, no. 4, pp. 30–46, 1987.
[12] D. Cope, *EMI: Experiments in musical intelligence*. A-R Editions, 1996.
[13] J. a. Biles, "GenJam: A genetic algorithm for generating jazz solos," *Proc. Int. Comput. Music Conf.*, pp. 131–137, 1994.
[14] N. V. Woods and N. V. Woods, "No Title," no. May, pp. 1–58, 2009.
[15] C. Turner, "Liquid Brain Music," University of Hull, 2008.
[16] G. Diaz-Jerez, "Composing with Melomics: Delving into the computational world for musical inspiration," *Leonardo Music J.*, vol. 21, pp. 13–14, 2011.
[17] T. Anders, "Composing Music by Composing Rules: Design and Usage of a Generic Music Constraint System," *Soc. Sci.*, vol. 10, no. February, p. 412, 2007.
[18] C. Ariza, "The interrogator as critic: The Turing test and the evaluation of generative music systems.," *Comput. Music J.*, vol. 33, no. 2, pp. 48–70, 2009.
[19] M. Y. Lo, "Evolving Cellular Automata for Music Composition with Trainable Fitness Functions," *Electron. Eng.*, no. March, 2012.
[20] M. T. Pearce and G. a Wiggins, "Towards A Framework for the Evaluation of Machine Compositions," *AISB Symp. AI Creat. Arts Sci.*, pp. 22–32, 2001.
[21] G. Ritchie, "Some Empirical Criteria for Attributing Creativity to a Computer Program," *Minds Mach.*, vol. 17, no. 1, pp. 67–99, 2007.
[22] M. a. Boden, "Computer models of creativity,"

Psychologist, vol. 13, no. 2, pp. 72–76, 2000.
[23] B. Manaris, J. Romero, P. Machado, D. Krehbiel, T. Hirzel, W. Pharr, and R. B. Davis, "Zipf's Law, Music Classification, and Aesthetics," *Comput. Music J.*, vol. 29, no. 1, pp. 55–69, 2005.
[24] S. Arshi and D. N. Davis, "Towards a Fitness Function for Musicality using LPM," *6th York Dr. Symp.*, 2015.
[25] A. W. Burks, "Von Neumann's Self-Reproducing Automata," *Essays on Cellular Automata*. pp. 3–64, 1970.
[26] A. Burks, *Essays on Cellular Automata*. University of Illinois Press, 1970.
[27] S. Wolfram, *A New Kind of Science*. 2002.
[28] D. Burraston and E. Edmonds, "Cellular automata in generative electronic music and sonic art: a historical and technical review," *Digit. Creat.*, vol. 16, no. 3, pp. 165–185, 2005.
[29] D. Burraston, E. Edmonds, D. Livingstone, and E. R. Miranda, "Cellular Automata in MIDI based Computer Music," *Proc. Int. Comput. Music Conf.*, vol. 4, pp. 71–78, 2004.
[30] E. R. Miranda, "Evolving cellular automata music: From sound synthesis to composition," *Proc. Work. Artif. Life Model. Music. Appl.*, p. 12, 2001.
[31] E. R. Miranda, "Sounds of artificial life," in *Proceedings of the 4th Conference on Creativity & and Cognition, ACM*, 2002, pp. 173–177.
[32] P. Hoffmann, "Towards an Automated Art: Algorithmic process in Xenakis' Composition," *Contemp. Music Rev.*, vol. 21, no. 2–3, pp. 121–131, 2002.
[33] E. R. Miranda, "Cellular automata music: From sound synthesis to musical forms," *Evol. Comput. Music*, pp. 170–193, 2007.
[34] C. Darwin, "The Origins of Species," in *Chapter 4, Natural selection*, London, England: OUP Oxford; Revised edition edition (13 Nov. 2008), 1906.
[35] B. P. Buckles and F. Petry, *Genetic Algorithms*. 1992.
[36] a R. Burton and T. Vladimirova, "Generation of musical sequences with genetic techniques," *Comput. Music J.*, vol. 23, no. 4, pp. 59–73, 1999.
[37] D. E. Goldberg, *Genetic Algorithm in search, optimization, and machine learning.* 1989.
[38] A. Horner and D. E. Goldberg, "Genetic algorithms and computer-assisted music composition," in *In Proceedings of the International Conference on Genetic Algorithms*, 1991.
[39] B. L. Jacob, "Composing with genetic algorithms," *Proc. 1995 Int. Comput. Music Conf.*, no. September, pp. 452–455, 1995.
[40] Biles, "GenJam," 2014. [Online]. Available: http://igm.rit.edu/~jabics/GenJam.html. [Accessed: 01-Oct-2014].
[41] P. Laine and M. Kuuskankare, "Genetic algorithms in musical style oriented generation," *First IEEE Conf. Evol. Comput.*, pp. 858–861, 1994.
[42] P. Dahlstedt, "Autonomous evolution of complete piano pieces and performances," *Proc. 9th Eur. Conf. Artif. Life*, no. 1999, 2007.
[43] M. Lo and S. M. Lucas, "Evolving Musical Sequences with N-Gram Based Trainable Fitness Functions," *2006 IEEE Int. Conf. Evol. Comput.*, no. 1, pp. pp. 601–608, 2006.
[44] R. Wooller and A. R. Brown, "Investigating morphing algorithms for generative music," *Procedings Third Iteration Third Int. Conf. Gener. Syst. Electron. Arts*, 2005.
[45] M. Farbood and B. Schoner, "Analysis and synthesis of Palestrina-style counterpoint using Markov Chains," in *In Proceedings of the International Computer Music Conference*, 2001.
[46] K. M. Biyikoglu, "A Markov Model for Chorale Harmonization," in *In Proceedings of the ESCOM Conference*, 2003.
[47] R. F. Voss and J. Clarke, "'1/f noise' in music: Music from 1/f noise," *Journal of the Acoustical Society of America*, vol. 63, no. 1. pp. 258–263, 1978.
[48] N. V. Woods, "Liquid Brain Music: Phase II," no. May, 2009.
[49] P. R. Cook and G. P. Scavone, "The Synthesis ToolKit in C++ (STK)," no. 09/10/2008. 2008.
[50] W. Li and P. Norman, "The Structure of the Elementary Cellular Automata Rule Space," *Complex Syst.*, vol. 4, pp. 281–297, 1990.
[51] D. N. Davis, "Computer and Artificial Music: Liquid Persian Music." [Online]. Available: http://www2.dcs.hull.ac.uk/NEAT/dnd/music/lpm.htm l. [Accessed: 01-Sep-2015].
[52] E. Jen, "Global properties of cellular automata," *J. Stat. Phys.*, vol. 43, no. October, pp. 219–242, 1986.

A Component-Based Architecture for Suspense Modelling

Pablo Delatorre[1] and **Barbara Arfè**[2] and **Pablo Gervás**[3] and **Manuel Palomo-Duarte**[4]

Abstract. Suspense is a key narrative issue in terms of emotional gratification, influencing directly the way in which the audience experiences a story. Disciplines like psychology, neurology or e-learning study the suspense as the basis of useful techniques for the treatment of mental diseases or improving memory skills and the comprehension. In the field of creativity, it's an essential cross strategy found in almost any book, film and video-game plots, regardless of technology and genre. With the objective of generating engaging stories, some automatic storytelling systems implement a suspense generation module. These systems are mainly based on narrative theories. However, we observe a lack of aspects from behavioral sciences, involving the study of empathy and emotional effect of scene objects in the audience. Generated plots with an adequate treatment of these features may involve benefits in areas as education and psychology. In this paper, we propose a component-based architectural model that firstly aims to identify and extract all these individual factors of the suspense from a scene; in a second step, the system calculates the level of suspense using a weighted corpus; in the last step, it alters those elements to increase or decrease the original suspense level and reassembles them in a new scene. Further, we discuss the model facing the development challenges and its practical implications.

1 INTRODUCTION

Nowadays, technology allows to make interactive or pseudo-interactive spaces of choice that were unthinkable just two decades ago. Not only there are new interfaces for innovative functionalities, but people have access to a practically infinite cosmos of potential multimedia experiences through which can choice what and how to start to consume, and when to stop it. Electronic books, films, serials, comics, music and web site written contents are within reach. If any choice is not good enough, other stimulus can be easily found. Even in case that the election has been satisfactory, the ease to move forward the broadcast content allows that superfluous parts can be missed, so, keeping the interest is essential to maintain the fidelity.

In this context of endless offers, the audience has become more demanding. An example is the visible decrease of the consumption of horror and suspense literature and movies. Statistics about the number of tickets sold in horror films are determinant: in 2014, 63% less tickets were sold than eight years ago, in 2006, following a downward arc from then[5]. Thriller genre seems to do better, with only a drop of 44% from 2013 to 2014 [1]. Therefore, the audience's score of the most watched horror movies per year conveys another important and continuous[6] descent of the assessment of the genre, from 71 points over 100 in 2006 ("Saw III") to 39 last year ("Annabelle") [2]. Blaming Internet piracy is not possible, while these statistics start in 2006, when piracy was already an established practice. In fact, even missing a serious official study from 2012, BSA Global Software Piracy Study reflects that Internet piracy rating globally decreased from 2006 to 2011, two points in North America and European Union [13]. More decisively, copycat products, predictability, annoying, boring, lack of atmosphere and insipid characters [51] affect the suspense and fear in a negative way. This seems to be leading to the apparently progressive decline of the most significant entertainment industry of suspense and fear during the last years [8], as happened with the horror literature in the nineties [28].

Suspense is an key narrative issue in terms of emotional gratifications. Reactions in response to this type of entertainment are positively related to enjoyment [40, p. 315], having a big impact on the audience's immersion and suspension of disbelief [29, p. 1359]. There is an interaction effect of negative valence and liking [4, p. 2]. The general pattern indicates that readers find literary texts interesting when the content is suspenseful, coherent, and thematically complex, accounting for approximately 54% of the variance in situational interest, where suspense made the single greatest contribution, explaining roughly 34% of variation [50, p. 445]. Consistent with this, experiments in video-games industry conclude that players find suspenseful games versions more enjoyable than non-suspenseful ones [34, p. 31].

Furthermore, suspense enjoyment is not subscribed only to the field of entertainment. For example, in the area of education, it is a direct way to create emotions that stimulate affective content, which influences positively the performance of the implicit and explicit memory [12, p. 223], and physiological responses in the way that we assimilate information [24]. Besides, in terms of psychological treatment, suspense is a subject of interest. To anticipating events as part of the experience of suspense is a creative problem solving that helps to counteract negative and stressful effects [58, p. 48]. Conversely, many people experience dramatic psychological consequences when they are exposed to suspenseful and fearful texts or movies. These consequences can be quite significant and include nightmares, physical stress, lingering fear, an increased heart rate, and heightened phobias [47, p. 48].

According with these arguments, we aim to adequate the regu-

[1] University of Cadiz, Spain, email: pablo.delatorre@uca.es
[2] University of Padova, Italy, email: barbara.arfe@unipd.it
[3] University Complutense of Madrid, Spain, email: pgervas@sip.ucm.es
[4] University of Cadiz, Spain, email: manuel.palomo@uca.es

[5] Two exceptions have been found: 2009 and 2013, supported only by the films "Paranormal activity" and "The conjuring" with almost fifteen millions of tickets sold each.

[6] One more time, the only exception has been "The conjuring", which in 2013 got 82 over 100.

lation of suspense as the one of the primary factor influencing the interest of a story as well as its benefits in other several fields. We support that suspense is a relevant cause of the audience's fearless and high emotional arousal level, and it has an important impact in the audience's rejection of this sort of exposition [47, p. 48]. In this context, our main goal is to propose a computer-based model which allows to increase or decrease on request this level of arousal in any potential suspenseful content, so it can enjoy with the same emotions under a different intensity anxiety level.

In this paper we explore a proposal to enrich a narrative by a computational model that makes possible two main goals: a) to measure the potential suspense intensity of any individual plot scene; b) to interactively increase or decrease that intensity by modifying the information that generates suspense, in a way that the suspense changes dynamically accordingly. In the section 2 we present the concept of suspense and its features. In section 3 we review several proposal systems that model suspense as element to generate stories. In section 4 we explain the model and detail the software architecture of our aimed application. Then, we finally discuss our proposal in section 5, and present our conclusions and argue about our future work in section 6.

2 ELEMENTS OF THE SUSPENSE AND READER AFFECTION

In this work, we adopt the Zillmann's definition of suspense [60, p. 208]: "A noxious affective reaction that characteristically derives from the respondents' acute, fearful apprehension about deplorable events that threaten liked protagonists, this apprehension being mediated by high but not complete subjective certainty about the occurrence of the anticipated deplorable event". This can be refined through the idea of suspense in drama as subject to further unique and seemingly universal restrictions: 1) the preoccupation with feared outcomes; 2) the selection of liked protagonists as targets for feared outcomes; and 3) the creation of high degrees of subjective certainty for the feared outcomes that threaten liked protagonists [59, p. 135].

In addition to this, we have identified three dimensions that essentially compose the suspense: emotional valence, empathy and arousal.

Emotional valence describes the extent to which something cause a positive or a negative emotion [19, p. 79]. In terms of the story, an element has a negative valence when it push towards a negative outcome. It has been extensively investigated the paradox in that texts with negative valence are perceived as more amusing than texts with neutral or positive valence. Citing Altmann [4, p. 2], the emotional reaction to uncomfortable expositions has been studied in media psychology regarding different narrative contexts. These contexts include tragic television news and crime drama [61][44][45], where enjoyment of unpleasant stories is not limited to a happy endings [49]. Suspense increases while the negative outcome probability [31, p. 107, 137] and the negative valence effect of the environment features do [26, p. 19].

According with this and as our second dimension, it has been tested that increased film enjoyment was reported for viewers with high empathy [22, p. 91]. Empathy is an emotional involvement related to the capacity to understand another's affective state [4, p. 2]. This positive emotional feeling occurs and increases with: a) the character's physical attractiveness [33, p. 2] and b) the audience's endorsement with the character's ethical behaviour and moral judgement [55, p. 344]. Specifically, suspense in fiction occurs, in general terms, when all likely outcomes are such that the outcome considered as correct is perceived to be much less probable [15, p. 137].

Generally, a seemingly effective manner to predict the level of empathy is through a measure of similarity and dissimilarity [44, p. 405]: a reader will imagine a preferred outcome for a character that the reader identifies with [16, p. 13]. Suspense will increase the further the story keeps away from that desired result. By throwing obstacles in the character way, the narrative can generate the anticipatory emotion of the fear [32, p. 7]. Although nowadays in the field of psychology it's not clear how many features may provoke empathy, some characteristics like race [18] and attractiveness [23] has been verified as generators of it.

Our third dimension, is the arousal [7], that refers the intensity of the emotion [19, p. 79]. This dimension seems to have a similar effect on the audience that the pattern found in negative valence. So, the higher the discomfort during the tension phase, the higher the pleasure in the moment of resolution [36, p. 82]. Novelists and narratologists agree with that the duration of the scene has an important role in this tension. "Suspense" comes from the world "suspend". Its etymology suggest that the more suspense is wanted, the longer suspend the scene is needed [39, p. 106]. Presenting the outcome a little later than expected [21, p. 325] is a key that relates suspense and timing.

Even considering that exists a possible dependency among these three dimensions (and more typically between valence and arousal) [10][29][20], we consider them separately. The reason is because, even if changes in one can affect the others with respect to the reader perception, we consider characteristics that can not be derived or transferred from one dimension to any other. For example, empathic features as victim physical attractiveness don't depend on negative valence nor the arousal, and cannot be extrapolated. Likewise, several models of emotion assume that valence and arousal are distinct variables [20, p. 324].

So far we have argued the dimensions of suspense in the narrative, we cannot obviate the variability of the effects in the individual spectators. It may be pointed out that emotional responses to narratives always are based some degree on personal experiences, real or experienced through fiction [56, p. 971], so, the suspense will be enjoyable or not depending of its impact in the reader. In addition, there is a fairly generalised consensus that this emotional effect is not just due to the story implicit or explicit characteristics (like the general meaning of words used), but the disposition of the reader to the patterns of the text [30, p. 279]. Under this standpoint, reading process is a reciprocal and transactional relation between the reader and the text. As the emotional effects, the "meaning" does not reside ready-made in the text or in the reader, but happens during the transaction between them [48, p. 4]. Without leaving behind the unquestionable relevance of the words, there is a broad agreement that the meaning created when reader and author meet in the texts is higher than written text or previous reader's knowledges [9, p. 31]. Thus, the power of the worlds is tempered to the reader's internal process of assimilation in its context where they are presented. However, there are common narrative features involved globally in suspense, although the level of emotional effects depend on individual characteristics of the audience.

This approach must be taken into account when we consider the three dimensions of the suspense. Therefore, we don't understand valence, empathy and arousal of a concept regardless of the reader or the viewer. Moreover, the modulatory effect of emotional context was constrained by the inherent meaning of target word [37, p. 379], of an image or both as significant [57, p. 404]. It has also been demonstrated that affective content affects men and woman differently [12, p. 219, 223]. Individual perceptions as the prediction of the outcome of events (including the character's fate) depends of the culture and upbringing [35, p. 124].

We can conclude that the individual estimation (weight) of the concepts like meaningful context are necessary to understand the individual suspense in their three dimensions: emotional valence, empathy and arousal.

3 AUTOMATIC SUSPENSE GENERATION

The treatment of suspense in the main narrative automatic systems is briefly summarized and discussed below.

MEXICA [42] is a program that writes short stories about the Mexicas, the old inhabitants of what today is Mexico City [42, p. 2]. These stories are represented as clusters of emotional links and tensions between characters, progressing during development, and whose operators, intensity and predefined texts are customizable. In MEXICA, it is assumed that a story is interesting when it includes degradation-improvement processes (ie., conflict and resolution) [42, p. 4]. Throughout the history, emotional links among the characters vary as a result of their interactions; so, *princess healed jaguar knight* produces the effect of increasing a positive emotion (gratitude) from the knight to the princess.

MEXICA is an exception in the use of positive emotions to implement the narrative tension. The system works with two predefined types of emotion: brotherly love and amorous love, both ranging from -3 (negative emotion) to 3 (positive emotion). Aditionally, ten types of tension are defined (actor dead, love competition, health normal...), which are generated based on the type and emotional value of each character. The stories search degradation-improvement curves through actions that transform the extent of the tensions.

MINSTREL [54], meanwhile, is a complex program that writes short stories about Arthurian legends, implemented on a case-based problem-solver where past cases are stored in an episodic memory [43, p. 4]. MINSTREL recognizes narrative tension plots and tries to increase the suspense by adding more emotionally charged scenes, by storing a simple ranking which tells when such inclusion is reasonable; for example, when the action is preserving a life. It uses two strategies for generating suspense: via character emotion and via character escape. In the first one, it is included in the text a sentence that describes the fear of the character about the immediate threat. The second one adds another sentence that reports a failed character's attempt [54, p. 123–126].

Another initiative is Suspenser [17], that adapts the cognitive psychology theories for creating stories with the objective of increasing the reader's suspense. It provides an intermediate layer between the fabula generation and the discourse generation, which selects the steps of the plot according to their *value of importance* for the final goal. For this and based on the Gerrig & Bernardo's assumption[7], Suspenser uses a set of heuristics grounded in the number of paths available for the character to reach its goal, considering optimal the probability of protagonists' success as 1/100 [16, p. 59].

Also based in Gerrig & Bernardo's work, Dramatis proposes an implementation of a system to evaluate suspense in stories that utilizes a memory model and a goal selection process [41, p. 5], assuming that the reader, when faced with a narrative, evaluates the set of possible future states in order to find the best option for the protagonist. With a similar target, Dramatis generates escape plans attempting to "break" the causal links that would reach non-desired goals (typically, the character death) and the reader could predict more easily. To do this, the memory model assigns more relevance to the elements recently narrated than to those mentioned at the beginning of the story.

Finally, we review IDtension [53], a drama project which comes up in order to demonstrate the possibility of combining narrative and interactivity. Unlike approaches based in character's chances or the course of the actions, it conceives the stories based on narrative properties (conflict or suspense).

Suspense is treated by IDtension as a reaction to the *obstacles* (conflicts), and is correlated to the risk of facing every expected obstacle (high or low risk, without intermediate values). The narrative effects of the tension are calculated by six criteria: ethical consistency, motivational consistency, relevance, cognitive load (influence in the story), characterization and conflict. Also, the condition is managed by a series of actions as accepting, refusing, congratulate, etc., available for use on / among the characters.

Although their purpose is not the generation of suspense, other proposals include mechanism for its treatment. IPOCL [46] is a kind of POCL planner (*Partial Order Causal Link*) which is improved by a factor of intention (*Intent-Driven*) on the part of the characters, denoted by a function `intends`; the planner attempts to satisfy the intention whenever possible. Indexter [14], based on IPOCL, offers a model that emulates the reader's memory, allowing use of knowledge previously presented as new (and surprising) information. To end this section, we mention Prevoyant [6], that enriches the stories with flashback (past events) and foreshadowing (hints of what is to come) strategies, providing additional data and tension ahead, respectively. For the first strategy, Prevoyand identifies causal links previous to the current event that haven't been described; as to the second, it submits a character or object participant of a future event.

With respect to our goals, the review of the above systems has exposed some comparative limitations. Firstly, we can observe that none of them takes into account the empathy as explicit part of the model, neither physical aspect nor moral and ethical issues. Additionally, as result of the evident limits of the current storytelling systems, the number of possible interactions between the characters and the environment is restricted by its respective internal base of events. In any case, MEXICA allows to redefine the actions.

On the other hand, all of them include arousal to a greater or lesser degree. MEXICA implements an emotional gradual intensity, ranging from -3 to 3, being the only that includes emotional valence both in this range and in the variability of possible interactions. MINSTREL, Suspenser, Dramatis and IDTension don't include the valence either. With respect to IDTension we haven't studied the effect of the interactivity narrative in the result of the suspense.

4 PROPOSED ARCHITECTURE

The main objective of the proposed architecture is the adaptation of the descriptive elements of a scene, in such a way that the scene output is adjusted to the required suspense intensity. Both the order and the result of the events will not be changed from the scene input to the output.

Following, we present our proposal. Figure 1 illustrates a simplified architecture of our model. This architecture consists in seven main elements (*Scene*, *Intensity*, *Extractor*, *Components transformer*, *Corpus*, *Reassembler* and *Output scene*) whose functions are described next.

[7] "Readers feel suspense when led to believe that the quantity or quality of path through the hero's problem space has become diminished". [27]

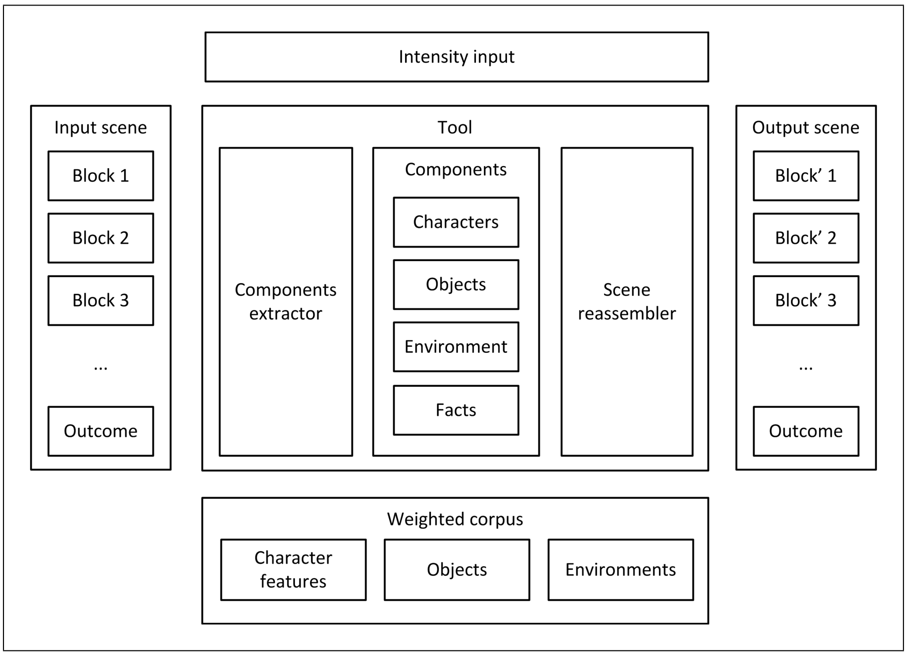


Figure 1. Proposed model architecture

4.1 Scene

Our proposal adopt the concept of *Scene* from the cinematographic sphere. A scene is the unity of dramatic action which, endowed with initial approach, junction and outcome, is determined by a spacial location criteria and change at least one of the values of any character's life[8] [5, p. 195][38, p. 56]. For our purposes, we consider a scene as a succession of information blocks that are provided to the audience. Such information blocks will be divided in descriptions and actions. Descriptions are representations of characters or objects, relating or explaining their different parts, features or circumstances [25]. These representations can be described by specific sentences or through the optional enrichment with adjectives of the part of the narrative where the element is referred. Meanwhile, actions constitute a succession of events during the scene.

We illustrate a fragment of the original script of Psycho film's shower scene [52] as example of these descriptions and actions:

Over the bar on which hangs the shower curtain, we can see the bathroom door, not entirely closed. For a moment we watch Mary as she washes and soaps herself.
There is still a small worry in her eyes, but generally she looks somewhat relieved.
Now we see the bathroom door being pushed slowly open.
The noise of the shower drowns out any sound. The door is then slowly and carefully closed.
And we see the shadow of a woman fall across the shower curtain. Mary's back is turned to the curtain. The white brightness of the bathroom is almost blinding.
Suddenly we see the hand reach up, grasp the shower curtain, rip it aside.

[8] "The scene is the unity of any dramatic action that, endowed with a beginning, middle and end, is determined by a criterion of spatial location" [5, p. 195]. "A scene is an action that occurs through a conflict in time and space more or less continuous, changing at least one of the values of the character's life" [38, p. 56].

In this fragment, we can observe the descriptions (bar, curtain, semi-closed door, Mary, sightly worried eyes, noise, shadow of a woman, white brightness and hand) and the actions (Mary washes and soaps herself, door is pushed slowly open, door is slowly and carefully closed and the shower curtain is grasped and ripped).

We represent each of the elements as an information block in the same order that they are narrated, being descriptions or actions. In any case, in this paper we focus on descriptions, as being the items which our model operates with. For this modeling, we are analyzing the description mode of some existent storytellings, searching for enough level of detail and robust knowledge representation. We put aside the process of actions for a future work.

4.2 Intensity

Intensity is represented as a quantitative integer value that indicates the desired level of suspense in the scene output. This way, if the difference between the required intensity and the intensity of the scene input is positive, the model will generate a more suspenseful scene, and vice versa. In case of coincidence, the scene output could involve a descriptive variation with respect to the scene input, but holding an equivalent suspense level.

As the intensity is an internal value whose range is not known by the user, we propose to specify the increment or decrement of the current intensity in the input. However, it is not necessary to include this option in the present model description, because the architecture is focuses on the functionality, not on the usability.

4.3 Extractor

The main process consists on three stages. In the first one, the extraction, description items have to be obtained from the information blocks. This step will depend on the scene input format: a) if scene input contains the states of the plot including specify tags for identifying the suspense elements, the extraction is immediate; b) if scene input is a narrative discourse, a complex algorithm will be required for analysis, identification and classification. We are working on the first option, leaving the discourse for later.

4.4 Components transformer

As we have referred, an intrinsic part of the extraction system is the transformation of the concepts. This component selects elements of the plot considering the interesting elements: characters, environment descriptors, object y facts. Among these, a second stage will select those elements implied in the plot according to their arousal and value of valence as measure of suspense affection.

After this, the goal of the transformation stage is to modify the descriptive elements, so the total amount of estimated suspense matches the intensity input. This implies the selection and launching of actions as Substitution, Insertion and Elimination (SIE) of characters features, characteristics of the environment and object's descriptions.

Transformation stage is complex: it implies making up a set of SIEs and select the most accurate in one way that the story keeps its consistency and the descriptor density deviation fits a minimum valid range. An evolutionary algorithm with customized heuristic of fitness is currently under production to satisfy our requirements.

4.5 Corpus

Our corpus consists on a set of terms, each one associated to a quantitative value that represents its level of suspense based on the already referred emotional valence, potential empathy and arousal. This information allows to measure the intensity of the scene from the previously extracted methods, in the transformation stage.

In the current state of our research, we base the corpus information in three big groups: character features, objects and environments.

4.5.1 Character features

As the same way than corpus is weighted using emotional valence, potential empathy and arousal, character features are balanced. According to several definitions of suspense, in terms of the characters the emotion generated is related (although not only) to the fate of at least one of them (victim) due to the actions of another one (threat). Directly or indirectly, this figure and its features are the center of what happens in a scene. Regarding Zillmann, the more dangerous and near is the threat, the more apprehension is experienced at the approach of the deplorable event. However, even this may seem obvious, we can not forget that the threat is not a static independent actor. On the contrary, the circumstances of the potential victim can change its nature.

For example, just before the mentioned Psycho's shower scene, the film script describes Mary's circumstances as [52]:

> *She goes into the bathroom, drops the pieces into the toiler bowl, flushes the toilet. Then she drops her robe and steps into the tub and turns the shower on.*
> *Over the bar on which hangs the shower curtain, we can see the bathroom door, not entirely closed. For a moment we watch Mary as she washes and soaps herself.*

This way, the screenplay is preparing the viewer for amplifying the effect of the immediate scene of suspense through the victim's features, portraying her as a helpless person: she is nude, can not hear or see well due to the shower, and her ability to escape is obviously limited[9]. On the contrary, if Mary had been warned, with the curtains open and a gun in her hands, suspense arousal would have been different comparing to the original script. Moreover, even if this is not the probable case, an extreme lack of sympathy for Mary could provoke that the spectator wouldn't really matter what will happen to her.

Summary up, a suspense generator must firstly take into account the character's features in the context of the threat, while the effect of its proximity in the viewer depends on the circumstances and characteristics of the victim that can be perceived and interpreted by the audience.

We propose three features to conform our corpus' *Character features* block: a) features related to balance of outcome oriented implicit strengths; b) features related to the empathy; and c) features related to the proximity between threat and victim to the outcome, as a spatial or temporal dimension concerning both sides.

The first one refers to the perceptible ability of the victim to counteract the threat. It implies an extensive ontology of features that includes physical aspects such as size, physical strength, intelligence, perceptive skills or endurance; capabilities such as experience in the use of weapon or fantastic abilities as crossing through walls; and resources at hand such as armours, guns or flashlights. These features are measured and quantified for both victim and threat, and the difference between both values represents the influence of the strength in the suspense. Considering the revisions about the suspense, our hypothesis is that the stronger the threat is for the victim, the more potential suspense can be generated.

The second character feature is about empathy: the features that generates feelings of identification in the viewer. As explained in the previous section, just few features have been proved as *empathizables*. Concretely, race and attractiveness[10] are two verified features that fit in the context of suspense. As well as the race can be easily represented by the model, the concept of "attractiveness" involves physical, mental, behavioural or derived from a position of power (including helplessness) which require a more complex treatment.

The last character feature refers to the proximity to the outcome. This is directly related to the scene *tempo*; as we have already mentioned, timing is an important criteria for evoking suspense. Therefore, suspense will increase as the threat is approaching the victim, physical (the killer) or just on time[11] (a countdown explosion). The behaviour of the victim and the threat are opposed: as the threat tries to reach the negative outcome, the victim struggles to get away from it. We consider that more suspense is inoculate in the audience the lower sum of the quantification of these distances is.

4.5.2 Objects

Objects involved in a suspense scene can take the role of: a) elements which influences the scene plot; b) decorative elements without direct participation. The model works on a different way depending on the case. As it depends on the context, this difference is not specified directly in the corpus.

[9] The effect of the escape's ways in suspense is already mentioned in Gerrig & Bernardo's assumption.

[10] Even the context in which its influence has been verified is very specific (a sexual aggression [23]), the consideration of that the feature can influence in the empathy of viewers in a suspense film is taken as hypothesis, extending the feeling to any kind of aggression as a generic helplessness situation.

[11] Physical approaching implies necessarily on time approaching, but not the reverse.

The elements that influences the plot are related to the character strength, while they can take the role of resources that can help or harm to the character. For example, in Psycho's shower scene, the curtain can be consider an element that creates a disadvantage for Mary, as the knife brings improvement for the killer objective. It is different from the balance of strengths of the block *Character features*: our concept of character features implies implicit attributes or *at hand* resources, but the *Object* block refers to potential resources. The effect in suspense is different: at the moment that the killer is approaching, it is not the same for the audience viewing the victim *having* a gun that realising that *there is* a gun on a table at hand (available for the first to catch it). The preferred objective and the expected steps change, influencing the perception of the proximity to the undesired outcome: defense capability of the victim is lower if the victim can not reach the weapon. Our proposal is that suspense is effective if the plot is pushed to balance the original difference of strengths while the outcome is approaching. Thus, the elements in the scene can contribute to strengthen or to weaken any of its parts.

The other kind of object has only a decorative function. Even though the influence of the aesthetic in objects with an active role in the plot, there are many others that just "colour" the localization. The valence of the elements influences its perception, which may have effect in suspense. For example, the toilet in Psycho's shower scene brings nothing to the events; moreover, it is not probably that the viewer can suppose any function related with the plot when the killer is approaching. However, Hitchcock decided to film it as he thought it would have emotional effects for the audience [3, p. 269].

4.5.3 Environments

Being called spacial context, atmosphere or scenery, the environment is a verifiable generator of suspense. In the one hand, it affects to the skills of the characters. Meteorological effects like fog and rain reduce the perception; snow makes the floor tricky; ice slides. As part of day cycle, twilight and night has similar impact in visual abilities. It affects to the balance of strength, usually negatively for the victim. For example, in Psycho's shower scene, the bathroom was full of steam; even if Mary had been facing the door or without curtains, it should have been hard to recognize the silhouette and to be on guard.

On the other hand, even if there is no objective reasons to have any kind of valence for an specific environment, we can not discard the classical conditioning: we have learnt that focusing a long corridor in an old castle usually precedes a negative outcome, even if there are no grounds to think that the corridor in and old castle is worse than a corridor in a beautiful mansion. This behaviour is similar to the decoration elements.

4.6 Reassembler

The scene reassembler is the part of the model responsible for building a new scened based on the original, putting all together the block of the plot, in the format of the chosen storytelling, as the has been modified by the Transformer step.

4.7 Output scene

On completion of the process, the scene output is the result of the model. We represents an example again from the scene input of the original script of Psycho film's shower scene [52], supposed a higher intensity required:

Over the bar on which hangs the shower curtain, we can see the bathroom door, not entirely closed. For a moment we watch Mary as she washes and soaps herself. ***Outside, we can hear a big storm in the middle of the night.***
There is still a small worry in her eyes, but generally she looks somewhat relieved. ***She feels guilty and sad.***
Now we see the bathroom door being pushed slowly open. ***A thunder resonates in the distance.***
The noise of the shower drowns out any sound. The door is then slowly and carefully closed.
And we see the shadow of a woman ***crawling toward*** *the shower curtain. Mary's back is turned to the curtain. The white brightness of the bathroom is almost blinding.*
Suddenly we see the hand reach up, grasp the shower curtain, rip it aside.

There are differences between the scene input and the result. Firstly, an environment description has been included (the storm). Besides, it is reported that Mary feels guilty and repentant. Finally, the shadow is weirder that in the original version.

5 DISCUSSION

Our proposed model does not only apply to movies, but to any kind of narrative created from a plot: games and automatic storytelling are good candidates to approach the system, as they can dynamically change part of the story before being presented to the audience. It provides the chance for changing the level of suspense as interaction with the story.

We do not expected problems in implementing technical issues, while nowadays there are plot generators, extractors and algorithms capable of generating different combinations of SIEs events according to a quantitative objective (for example, the scene arousal level). Once we get a functional model, a future objective will be extending the model beyond descriptions and characters features, but with facts (like movements of curtains or doors). However, our immediate challenge at this time is to get the *weighted* corpus, proposing a formal model for giving that quantitative measure for each element: within narratives a one-to-one mapping between words denoting emotions and actual experienced emotion is rarely found [56, p. 964].

At the time of this writing, we are working on it. The selection and analysis of suspense and horror movies may help to develop a first ontology and subsequent formulation of characters features; from the field of psychology, the study of classical conditioning comparing with the classical scenes will provide information to measure the importance of the environment in the arousal; finally, the revision of different studies of emotional affection generated by physical concepts will be useful to quantify the effect of the decoration objects [57][56].

Related to this, there are other limitations that we need to consider. We have distinguished between objects which influences the scene plot and decorative elements without direct participation. Assuming that some elements have an evident and natural effect in suspense because of its common purpose (knives, corpses or wardrobes), the utility of other specific objects in the plot is individually determined by each spectator or can depend on the context. Since the semantic meaning of a word can be selected by context, it seems reasonable that evaluation of the emotional tone of a word could be shaped by an emotional context, as emotional evaluation is more subjective and changeable than semantic meaning [37, p. 380]. For example, a laptop on a table can be only an inoffensive decoration, but it could be used as a blunt weapon too. It might lead to consider decorative effects negligible while we have other effects that influence directly the plot. We need conclude our study to confirm this hypothesis.

An additional difficult issue has been found analyzing the effect of the identification between audience and character. Although it is included in the model, we suspect that the concept of *empathy* (as an emotional, social and unconscious approach based on identification) has not a determinant weight in the suspense. This is extensible to other related aspects like moral or ethical behaviour. Therefore, we do not agree with a moral disagreement with a character is enough to desire an awful outcome, and a moral agreement is not enough to feel more suspense. We need to analyze the concept and effect of empathy in order to support this assertion.

At all events, we conceive the existence of an "internal conflict" due to the spectator's belief in a "just world" [11, p. 114, 116] that makes the vision of a character under threat as root of discomfort. However, this feeling seems independent of the empathy, as we sense it even the character actions had been at the antipodes of our ethical criteria.

Finally, we are convinced that the balance of implicit strengths and the proximity between threat and victim to the outcome are important features affecting to the suspense.

6 CONCLUSIONS

In this paper, we have presented a model schema that aims to: a) measure the arousal of a suspense scene; b) compose a new scene by replacing and adding elements and characters features, adapting the audience's preferred intensity level. We consider this objective as interesting in terms of enjoyment: while some people experience excitement and intellectual stimulation when watching suspenseful or horror films, others experience dramatic psychological consequences. People more sensitive to this kind of emotional immersion are influenced to think that something bad is going to happen, and they report experiencing physical stress waiting for it to actually happen: being so scared that they are afraid to go home after the movie or walk to their cars in the parking lot, or need to sleep with someone else. They get truly scared and are affected physically and psychologically [47]. Providing to this people to enjoy and share horror common films is a reason for measuring and adjust the suspense arousal.

Besides, in the field of the computational creativity we expect that this general quantitative prediction model will serve as a basis for benchmarks on stories based on their potential interest to the viewer, in the form of suspense. Likewise we aim to provide, for automatic, interactive or supervised storytelling generators, models of decision regarding choosing conceptual spaces in the plot development.

Although the model is defined enough, some specifications must still be concreted. Firstly, potential storytellers, one of which will support the descriptions of the plot, objects and characters features, are currently being analyzed. Secondly, we need explore more deeply the relation between suspense and its dimensional components: emotional valence, empathy and arousal. This will allow us to determine a quantitative formal method to assign the weight to the elements involved in the stories. Finally, an optimal algorithm is under study to select the most adequate SIEs depending on the required intensity.

ACKNOWLEDGEMENTS

This work has been funded by the Andalusian Government under the University of Cadiz programme for Researching and Innovation in Education. This paper has been partially supported by the projects WHIM 611560 and PROSECCO 600653 funded by the European Commission, Framework Program 7, the ICT theme, and the Future and Emerging Technologies FET program.

REFERENCES

[1] The Numbers©. Nash Information Services, LLC, `http://www.the-numbers.com/`, 2015. accessed on 6 January 2016.
[2] Rotten Tomatoes™. Flixster, Inc., `http://www.rottentomatoes.com/`, 2015. accessed on 6 January 2016.
[3] Richard Allen and Sam Ishii-Gonzales, *Hitchcock: Past and Future*.
[4] Ulrike Altmann, Isabel C Bohrn, Oliver Lubrich, Winfried Menninghaus, and Arthur M Jacobs, 'The power of emotional valence—from cognitive to affective processes in reading', *Frontiers in human neuroscience*, **6**, (2012).
[5] Daniel Aranda and Fernando De Felipe, *Guión audiovisual*, Editorial UOC, 2006.
[6] Byung-Chull Bae and R Michael Young, 'A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach', in *Interactive Storytelling*, 156–167, Springer, (2008).
[7] Daniel E Berlyne, 'Conflict, arousal, and curiosity.', *McGraw-Hill Series in Psychology*, **XII**, (1960).
[8] Alex Billington. The decline of the horror genre? First Showing, LLC, April 8th, `http://www.firstshowing.net/2008/sunday-discussion-the-decline-of-the-horror-genre/`, 2008. accessed on 6 January 2016.
[9] Esther Blanco Iglesias, 'La lectura de textos literarios. Una propuesta didáctica para la enseñanza de la literatura española a estudiantes brasileños', *Biblioteca Virtual redELE*, **2007**(7), (2005).
[10] Margaret M Bradley and Peter J Lang, 'Affective norms for english words (ANEW): Instruction manual and affective ratings', Technical report, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, (1999).
[11] William F Brewer, 'The nature of narrative suspense and the problem of rereading', *Suspense: Conceptualizations, theoretical analyses, and empirical explorations*, 107–127, (1996).
[12] Leslie A Burton, Laura Rabin, Susan Bernstein Vardy, Jonathan Frohlich, Gwinne Wyatt, Diana Dimitri, Shimon Constante, and Elan Guterman, 'Gender differences in implicit and explicit memory for affective passages', *Brain and Cognition*, **54**(3), 218–224, (2004).
[13] Bussiness Software Alliance, 'Shadow market: 2011 BSA global software piracy study', *The Compliance Gap*, (2012). `http://globalstudy.bsa.org/2011/downloads/study_pdf/2011_BSA_Piracy_Study-Standard.pdf`, accessed on 16 March 2015.
[14] Rogelio E Cardona-Rivera, Bradley A Cassell, Stephen G Ware, and R Michael Young, 'Indexter: A computational model of the event-indexing situation model for characterizing narratives', in *The Workshop on Computational Models of Narrative at the Language Resources and Evaluation Conference*, pp. 32–41, (2012).
[15] Noël Carroll et al., *The philosophy of horror: Or, paradoxes of the heart*, Routledge (New York), 1990.
[16] Yun Gyung Cheong, *A computational model of narrative generation of suspense*, Ph.D. dissertation, North Carolina State University, 2007.
[17] Yun-Gyung Cheong and R Michael Young, 'A computational model of narrative generation for suspense', in *Association for the Advancement of Artificial Intelligence (AAAI Journal)*, pp. 1906–1907, (2006).
[18] JY Chiao and VA Mathur, 'Intergroup empathy: how does race affect empathic neural responses?', *Current Biology (CB)*, **20**(11), R478, (2010).
[19] Francesca MM Citron, Marcus A Gray, Hugo D Critchley, Brendan S Weekes, and Evelyn C Ferstl, 'Emotional valence and arousal affect reading in an interactive way: neuroimaging evidence for an approach-withdrawal framework', *Neuropsychologia*, **56**, 79–89, (2014).
[20] Francesca MM Citron, Brendan S Weekes, and Evelyn C Ferstl, 'How are affective word ratings related to lexicosemantic properties? evidence from the sussex affective word list', *Applied Psycholinguist*, 1–19, (2012).
[21] Minet de Wied, Ed SH Tan, and Nico Henry Frijda, 'Duration experience under conditions of suspense in films', *NATO ASI series. Time, action and cognition: Towards bridging the gap*, 325–336, (1992).
[22] Minet De Wied, Dolf Zillmann, and Virginia Ordman, 'The role of empathic distress in the enjoyment of cinematic tragedy', *Poetics*, **23**(1), 91–106, (1995).

[23] Sheila R Deitz, Madeleine Littman, and Brenda J Bentley, 'Attribution of responsibility for rape: The influence of observer empathy, victim resistance, and victim attractiveness', *Sex Roles*, **10**(3-4), 261–280, (1984).

[24] Pablo Delatorre and Barbara Arfè, 'Modulare la suspense del lettore attraverso un modelo computazionale', in *XXVIII Congresso Nazionale Sezione di Psicologia dello sviluppo e dell'educazione, Parma (Italy)*, (2015).

[25] Real Academia Española, *Diccionario de la lengua española*, number v. 1 in Diccionario de la lengua española, Editorial Espasa Calpe, 2001.

[26] Jonathan Frome and Aaron Smuts, 'Helpless spectators: Generating suspense in videogames and film', *TEXT technology*, **13**, 13–34.

[27] Richard J Gerrig and Allan BI Bernardo, 'Readers as problem-solvers in the experience of suspense', *Poetics*, **22**(6), 459–472, (1994).

[28] Steffen Hantke, 'The decline of the literary horror market in the 1990s and Dell's Abyss series', *The Journal of Popular Culture*, **41**(1), 56–70, (2008).

[29] Chun-Ting Hsu, Markus Conrad, and Arthur M Jacobs, 'Fiction feelings in Harry Potter: haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience', *NeuroReport*, **25**(17), 1356–1361, (2014).

[30] Wolfgang Iser, 'The reading process: A phenomenological approach', *New literary history*, **3**(2), 279–299, (1972).

[31] Yumiko Iwata, *Creating Suspense and Surprise in Short Literary Fiction: A stylistic and narratological approach*, Ph.D. dissertation, University of Birmingham, 2009.

[32] Patrick Keating, 'Emotional curves and linear narratives', *The Velvet Light Trap*, **58**(1), 4–15, (2006).

[33] Jason Kelley, 'The effect of physical attractiveness on mirror neuron activity', *Summer Research*, **181**, (2013).

[34] Christoph Klimmt, Albert Rizzo, Peter Vorderer, Jan Koch, and Till Fischer, 'Experimental evidence for suspense as determinant of video game enjoyment', *CyberPsychology & Behavior*, **12**(1), 29–31, (2009).

[35] Caryn E Krakauer and Patrick H Winston, 'Story retrieval and comparison using concept patterns', *Proceedings of Computational Models of Narrative*, 119–124, (2012).

[36] Moritz Lehne, *Emotional Experiences of Tension and Suspense*, Ph.D. dissertation, Freie Universität Berlin, Germany, 2014.

[37] Hongyan Liu, Zhiguo Hu, and Danling Peng, 'Evaluating word in phrase: The modulation effect of emotional context on word comprehension', *Journal of psycholinguistic research*, **42**(4), 379–391, (2013).

[38] Robert McKee and Jessica Lockhart, *El guión. Story*, Alba Editorial, 2011.

[39] Raymond Obstfeld, *Novelist's Essential Guide to Crafting Scenes*, Novelists Essentials.

[40] Mary Beth Oliver, 'Exploring the paradox of the enjoyment of sad films', *Human Communication Research*, **19**(3), 315–342, (1993).

[41] Brian O'Neill, *A computational model of suspense for the augmentation of intelligent story generation*, Ph.D. dissertation, Georgia Institute of Technology, 2013.

[42] Rafael Pérez y Pérez, 'Employing emotions to drive plot generation in a computer-based storyteller', *Cognitive Systems Research*, **8**(2), 89–109, (2007).

[43] Rafael Pérez y Pérez and Mike Sharples, 'Three computer-based models of storytelling: BRUTUS, MINSTREL and MEXICA', *Knowledge-based systems*, **17**(1), 15–29, (2004).

[44] Arthur A Raney, 'Moral judgment as a predictor of enjoyment of crime drama', *Media Psychology*, **4**(4), 305–322, (2002).

[45] Arthur A Raney and Jennings Bryant, 'Moral judgment and crime drama: An integrated theory of enjoyment', *Journal of Communication*, **52**(2), 402–415, (2002).

[46] Mark O Riedl and R Michael Young, 'Narrative planning: balancing plot and character', *Journal of Artificial Intelligence Research*, **39**(1), 217–268, (2010).

[47] Tim Robinson, Clark Callahan, and Evans Keith, 'Why do we keep going back? A Q method analysis of our attraction to horror movies', *Operant Subjectivity: The International Journal of Q Methodology*, **37**(1-2), 47–57, (2014).

[48] Louise M Rosenblatt, *Writing and Reading: The Transactional Theory. Technical Report No. 416.*, ERIC, 1988.

[49] Holger Schramm and Werner Wirth, 'Exploring the paradox of sad-film enjoyment: The role of multiple appraisals and meta-appraisals', *Poetics*, **38**(3), 319–335, (2010).

[50] Gregory Schraw, Terri Flowerday, and Stephen Lehman, 'Increasing situational interest in the classroom', *Educational Psychology Review*, **13**(3), 211–224, (2001).

[51] Christopher D. Spicer. The decline of modern horror films: we need the scares back in 'scary' movies. Collective Publishing Company, Inc., October 23th, `http://www.thecollectivepc.com/2012/10/the-decline-of-modern-horror-films-we.html`, 2012. accessed on 6 January 2016.

[52] Joseph Stefano, *Psycho (script reviewed)*, 1959. Based on the novel by Robert Bloch.

[53] Nicolas Szilas, 'IDtension: a narrative engine for interactive drama', in *Proceedings of the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE) Conference*, volume 3, pp. 187–203, (2003).

[54] Scott R. Turner, *The Creative Process: A Computer Model of Storytelling and Creativity*, Taylor & Francis, 2014.

[55] Peter Vorderer, Silvia Knobloch, and Holger Schramm, 'Does entertainment suffer from interactivity? the impact of watching an interactive tv movie on viewers' experience of entertainment', *Media Psychology*, **3**(4), 343–363, (2001).

[56] Mikkel Wallentin, Andreas Højlund Nielsen, Peter Vuust, Anders Dohn, Andreas Roepstorff, and Torben Ellegaard Lund, 'Amygdala and heart rate variability responses from listening to emotionally intense parts of a story', *Neuroimage*, **58**(3), 963–973, (2011).

[57] Roel M Willems, Krien Clevis, and Peter Hagoort, 'Add a picture for suspense: neural correlates of the interaction between language and visual information in the perception of fear', *Social cognitive and affective neuroscience*, **6**(4), 404–416, (2011).

[58] Konstantinos Zachos and Neil Maiden, 'A computational model of analogical reasoning in dementia care', in *Proceedings of the Fourth International Conference on Computational Creativity*, pp. 48–55, (2013).

[59] Dolf Zillman, 'Anatomy of suspense', in *The entertainment functions of television*, 133–161, Psycology Press, (1980).

[60] Dolf Zillmann, 'The psychology of suspense in dramatic exposition', *Suspense: Conceptualizations, theoretical analyses, and empirical explorations*, 199–231, (1996).

[61] Dolf Zillmann, Kay Taylor, and Kelly Lewis, 'News as nonfiction theater: How dispositions toward the public cast of characters affect reactions', *Journal of Broadcasting & Electronic Media*, **42**(2), 153–169, (1998).