LRN2CRE8

LEARNING TO CREATE

PROJECT NUMBER: 610859

SMALL OR MEDIUM-SCALE FOCUSED RESEARCH PROJECT

ICT - FUTURE AND EMERGING TECHNOLOGIES (FET)

Deliverable 5.2:
Review of Methods

QMUL Version 1.0 FINAL

Executive Summary: This paper provides an overview of current methods used to assess human and machine creativity, as well as methods from other fields, such as psychology and neuroscience, which may be adapted for the study of computational creativity. Arguably, the field of computational creativity lacks systematic empirical evaluation of creative systems and their output. Motivation for more rigorous scientific testing is provided. Further, we highlight reasons why understanding human creativity can lend insight into computational approaches to creativity.

After providing leading theoretical and operational definitions of creativity, methods discussed in the remainder of the paper are divided into three broad categories: behavioral, cognitive, and neural. The behavioral section provides traditional methods for testing creativity in humans, such as divergent thinking tasks and the evaluation of artistic artefacts. The cognitive methods span a wide range of approaches, from conceptual spaces theory to cutting-edge techniques in machine learning. We take the theoretical stance that examining the creative process is just as important as evaluating a creative artefact. Therefore, emphasis is placed on learning and expectation mechanisms underlying knowledge representation and creative behaviours. And finally, the section focusing on neuroscientific approaches covers techniques such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) that may be used to assess the neural processes and functional organization of brain areas implicated in creative tasks.

The methods outlined in this paper are by no means exhaustive, but we hope to support an expansion of the current methods utilized for testing computational creativity, as well as some theoretical motivation for more rigorous empirical evaluation.

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	-
RE	Restricted to a group specified by the Consortium (including the Commission Services)	-
CO	Confidential, only for members of the Consortium (including the Commission Services)	-

Expanding the Empirical Evaluation of Computational Creativity: Theoretical Motivations and a Methodological Overview

Introduction

Creativity is a hugely important facet of humanity, and yet quite difficult to define. This poses a challenge to developing quantitative measures for assessing creative products and behaviours. Further, in the burgeoning field of computational creativity, there is the additional difficulty that the creative agents are not human, which can introduce biases in terms of evaluation. Although researchers in the field have occasionally agreed upon some general concepts and attributes of creativity – such as a confluence of *novelty*, *value or appropriateness* (of solution given a problem), and *fluency* – creativity must ultimately be assessed in terms of the particular domain being explored. In other words, the evaluation of creativity is domain and task dependent.

This paper aims to provide an overview of methods for creative evaluation, some developed explicitly for creativity and others adapted here for that purpose, drawing upon techniques from a diversity of fields. In order to provide explicit examples of tests for computational creativity, some focus will be placed on musical creativity. Computational methods provide many useful evaluation measures, but this paper argues that understanding human creativity is of great significance for developing and evaluating computationally creative systems. Computational evaluation methods may include self-referential techniques such as cross-validation, externally referential methods for testing the model's output against data or a theory, or evaluative judgements from a non-computational (human) examiner.

Knowledge of human creativity provides insight into ways of evaluating and analysing creativity. Probing human creativity, from the perspective of perception, cognition, attention, statistical learning, conceptual frameworks, etc, will help us understand the processes supporting creative behaviour. This, in turn, can help inform models of creativity, by suggesting approaches or algorithms, for example, which may facilitate implementation of the creative process in artificial systems. To provide one example, knowing that neural mechanisms exist which specifically respond to violation of expectation can provide biological grounding to novelty evaluation metrics, and bestow insight into which *kinds* of novelty are the most psychologically interesting and impactful.

Evaluating creativity often focuses around measures of novelty and value, which both depend upon the knowledge and state of the observer, and the context in which the creativity is observed. Indeed, as with any critique of art or creative output, evaluation of creativity depends on four factors: the creative artefact itself, the creator, the context in which the artefact is perceived, and the properties of the observer (that is, the evaluator's relevant experience, knowledge, and expectations). It is important to emphasize that creativity, and our assessment of creativity, is situated in a personal, social, and cultural framework (Csikszentmihalyi, 1999; Saunders & Gero, 2001).

Importantly, evaluation of creative systems should not be limited to solely the generated output. The agent's means of acquiring and storing information, manipulating and abstracting knowledge representations, and methods for achieving creative solutions are often just as important for assessing creativity as the generated output itself. When assessing a model's computational creativity, one can examine the *internal structure and processing dynamics* of the computational model, such as learned feature spaces or representations. Alternatively, one can examine the *generated output* or artefacts of the model, using for example evaluation metrics based on novelty and value, or emotional induction or physiological responses of humans to perception of creative versus mundane artefacts.

The field of computational creativity is largely lacking a formalization of techniques and more rigorous empirical testing of hypotheses, models, processing mechanisms and artifacts. A recent survey of methods used in computational creativity research confirmed that there is little rigorous scientific evaluation of these systems, but also discovered that relatively few papers even discuss (let alone test) whether their system is creative (Jordanous, 2012). Therefore, another important aim of this paper is to provide an overview of existing methods that may be applied to examining creative artefacts and creative processes. Rigorous empirical methods, with testable hypotheses and reproducible experimental designs, may allow disparate approaches within the field to converge upon key perspectives and objectives. More cohesion in the field could provide not only a methodological framework, but also a theoretical framework for interpreting incremental progress in the field of computational creativity.

Definitions of human and computational creativity

It is not the goal of this paper to enter the debate on defining creativity, but an overview of the various ways of defining and conceptualizing creativity is an essential background for speaking about creative evaluation. Settling upon one formal definition of creativity is not required here; rather, considering the alternative views and definitions simply produces a larger set of testable hypotheses, which may generate a broader set of evaluation tools.

One of the first attempts to address the process of creativity was by Graham Wallas. Wallas (1926) suggested that the creative process consists of four phases, preparation, incubation, illumination, and verification. Preparation involves the study and definition of a problem or task. Incubation is a time to let subconscious processes work by not placing explicit attention on the area for some time. Illumination is the emergence of a new idea in consciousness, and verification is the validation of the worth and value of this idea. Some years later, Koestler (1964) formulated a general theory of human creativity that introduced the idea of *bisociation*: a blending of ideas drawn from previously unrelated semantic matrices of thought. The theory focuses on abstraction, categorization, and analogy as a means of forming bisociations. This approach was very influential for future models of creativity, especially the theory of conceptual blending (Turner & Fauconnier, 1995; Fauconnier & Turner, 2002). Conceptual blending provides a framework and terminology for discussing creative thought processes, but does not address the ontology and origin of the ideas that are blended.

One of the first in-depth, theoretical discussions of computational creativity was by Margaret Boden. In her book *The Creative Mind* (1990), Boden identifies three different types of creativity: combinatorial, exploratory, and transformational. Combinatorial creativity is the process of amalgamating elements of existing concepts to form a new concept. For example, within a hypothetical conceptual space of possible architectural features (materials used, building style, functionality, etc), one can find a new combination of features to meet some goal, i.e. use reflective materials developed for space travel in order to enhance the sound quality of a performance space. Exploratory creativity involves discovering novel concepts or items within an existing conceptual space. And lastly, transformational creativity involves changing the structure or rules defining the conceptual space itself, or the type of traversal through this space. Transformational creativity thus produces a shift in thought or paradigm, enabling a new set of possible creative objects. Boden also offers the distinction between psychological creativity, or P-creativity, which refers to a particular individual's creativity, and historical creativity, or H-creativity, which is creativity recognized as novel at the societal level (Boden, 1990). Geraint Wiggins has formalized Boden's conceptual spaces theory in his *creative systems framework* (CSF) (Wiggins, 2006a). This has provided a shared terminology that enables discourse about creative systems (see for example, Ritchie, 2007, and Maher, 2010). Importantly, the CSF also specifies a framework for identifying the value of aberrant ideas. Additional ways of conceiving of and modelling conceptual spaces will be discussed in more depth below.

Further, Amabile (1983, 1985) has proposed the distinction between intrinsically and extrinsically motivated creativity. Intrinsic creativity is driven by internal reward systems, whereas extrinsic motivations arise from sources of external reward, such as recognition, alleviation of a problem or punishment, or financial gains. External motivation is argued to be detrimental to creativity, while intrinsic motivation can produce artefacts that are judged to be more creative by experts from the domain (Amabile, 1985).

The various views of human creativity are important for the investigation of computational creativity as well as its definition. As offered by Wiggins (2006), computational creativity is "The support, study and simulation, by computational means, of behaviours which would be deemed creative if exhibited by a human." The term "creative" is not defined more specifically because it refers to a subjective assessment based on cultural standards and preferences. However, to avoid the possible misconstrual that for something to be deemed creative, it must be of the type created by humans, this definition was later amended as follows: Computational creativity is "the philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative" (Colton & Wiggins, 2012). This operational definition also makes an explicit point of reducing creativity into constituent responsibilities or mechanisms, thereby providing another argument for the importance of evaluating creativity based on specific underlying properties and tasks.

Embracing this divide and conquer approach, Graeme Ritchie has proposed a set of criteria for assessing the output of a potentially creative system (note that he focuses on the artefact, and not on evaluating the internal processes underlying the creation of that artefact) (Ritchie, 2007). His empirical criteria are based on novelty (in terms of untypicality, innovation, or class membership) and quality (in terms of value ratings), and allow for variation in subjective judgement of creative artefacts as well as the variation in the criteria used to define creativity itself (Ritchie, 2001). .

Another framework for evaluating creativity that focuses both on generated artefacts and the behaviour that gives rise to those artefacts is Colton's *creative tripod* (Colton, 2008). Colton emphasizes that knowledge of the creative process influences an observer's judgements of creativity, in value judgements of human creativity, but also and especially in the context of computational systems. There is often a bias against artificial systems, where artefacts are judged to be less creative, or observers attribute the real creativity to the programmer. The creative tripod is a technique for describing and evaluating the behaviour of creative systems, and is based on assessing skill (technical ability), appreciation (valued in the domain), and imagination (appropriate novelty that moves beyond pastiche) (Colton, 2008). If the system is considered skillful, appreciative, and imaginative, then the software should be deemed creative. Colton and colleagues have also worked to begin formalizing computational creativity theory (CCT) by offering two descriptive models as a starting point: one focused on the act of creative generation, called FACE, and another which tests the impact that creative systems may have on the observer(s), called IDEA (Colton, Charnley, & Pease, 2011).

The remainder of this paper will focus on three main levels of analysis in understanding and evaluating creativity, namely the *Behavioral, Cognitive,* and *Neural* levels. Behavioural assessment concerns individuals' production of creative ideas or artefacts, or creative performance, as assessed at the behavioural level. The cognitive level concerns the acquisition and manipulation of concepts. This section forms the majority of this methodological overview, spanning perceptual processing, statistical mechanisms underlying expectation and learning, semantic/schematic representations, and approaches such as formal geometric models (e.g., of conceptual spaces), and symbolic symbol systems. Finally, the neural level focuses on electrophysiological, electromagnetic, and imaging techniques used to understand the neural basis of creativity.

Behavioural tests of creativity

Because we are addressing the evaluation of creativity, and creativity is a human construct, techniques for assessing creative processes and output should be applicable to both humans and machines. Given the long history of testing creativity in humans, we begin with an overview of behavioural tests of creativity before moving on to computational approaches.

Behavioural tests of creative thinking are fairly common, as assessing creativity is relevant for many professions, from army soldiers to CEOs. Sets of behavioural tests (called batteries) are also commonly used for children's entrance exams, identifying gifted students, or placement into enrichment classes or specialty art schools. Most behavioural batteries test divergent thinking, convergent thinking, artistic ability, or rely on self-assessment. Divergent thinking is the ability to generate new ideas or solutions to some problem or task. Because divergent thinking can result in non-novel ideas, originality, applicability, and value metrics are often employed in conjunction with divergent thinking tasks. Conversely, convergent thinking tasks measure the individual's ability to discover the one solution to a problem, and often require employing different strategies to find the solution. Self-assessment tasks prompt the individual to report his/her experience with creative pursuits, or ask questions about personality and creative inclinations. And lastly, artistic tests are usually domain-specific methods for testing proficiency or evaluating the creativity of an artistic object (such as a poem or musical score), often relying upon expert judgement.

The most common behavioural battery of creativity is the Torrance Tests of Creative Thinking, or TTCT (Torrance, 1974, and re-normed in 1974, 1984, 1990, and 1998), and may be employed for testing kindergarteners up through adults. The TTCT primarily tests divergent thinking ability, and is comprised of verbal and figural problems. Measures of assessment for the original TTCT were based on Guilford (1959), including Fluency, Flexibility, Originality, and Elaboration. Fluency measures the number of generated ideas that are relevant to the problem (such as the number of figural images produced in a drawing task). Flexibility measures the similarity of responses, and whether solutions are based on multiple domains or ways of thinking. Originality refers to the number of responses that are statistically infrequent. And lastly, Elaboration measures the individual's ability to develop and add ideas to pre-existing ones (Torrance, 1990; also see Kim, 2006).

When re-normed, two new measures were added, and Flexibility was eliminated because it was found to be highly correlated with Fluency (Kim, 2006). This yielded the five measures of Fluency, Originality, Elaboration, Abstractness of Titles, and Resistance to Premature Closure. Abstractness of Titles refers to the degree of understanding and abstraction of thought beyond mere labelling of pictures. Resistance to Premature Closure measures the amount of "psychological openness" used when considering alternative ideas and processing the available information (Kim, 2006). In addition to these five measures, Torrance (1990) identified a set of thirteen subscales of "creative strengths", including, for example, *richness of imagery*, *humor,* and *emotional expressiveness*.

Various other tests of divergent thinking, some of which were predecessors to the TTCT, are still utilized as well. Most of these tests examine individuals' elaboration of possible solutions, measuring detail, quantity, novelty, and variety of ideas. One approach by Wallach and Kogan (1965) has participants list as many items as possible that fall into a particular category. Another example is the Alternative Uses Task, developed by Guilford (1967), which asks participants to enumerate possible uses for a common object. There are also domain-specific divergent thinking tasks, which also fall under the category of artistic ability assessment. Many of these are also based on the work of Guilford and Torrance. The

Measures of Musical Divergent Production (Gorder, 1980), for example, tests musical participants on the number of improvised sounds produced (Fluency and Elaboration), shifts of musical content (Flexibility), Novelty, and "musical appeal". Because this assessment requires musical knowledge, scoring is performed by expert musicians. For other examples of musical creativity tests, see the Measure of Musical Problem Solving (Vold, 1986), and the Measure of Creative Thinking in Music II, (Webster, 1987).

Convergent thinking tasks approach creativity in a slightly different manner. To successfully complete these tasks, creative thinking is required in terms of reassessing a problem, finding an insightful problem-solving perspective, or shifting strategies to hone in on the correct solution. An example of a convergent thinking task is the Remote Associates Test (Mednick, 1962), in which participants must find one word that associates three given words.

Self-assessment tests implore the examinee to reflect on their own creativity, or report autobiographical information, such as the participant's involvement in the Arts (eg., experience taking a visual arts class or writing a short story). Again, many of these tests are based on the work of Torrance, such as the Khatena-Torrance Creative Perception Inventory (Khatena & Torrance, 1976). It should be noted that there is active discussion about the extent to which these tests examine creativity, as opposed to intelligence, or even personality traits such as individuality and self-confidence. Therefore, these tests may be slightly less pertinent to the field of computational creativity. That said, some work in the field aims to prompt computationally creative systems to examine and critique their own output (Colton, et al., 2014).

Another approach that has been successfully implemented for evaluation of creative artefacts is the Consensual Assessment Technique (CAT) (Amabile, 1982, 1996). In this approach, a panel of experts from the relevant field judge the creativity of an artwork, theory, etc. The method combines the critiques of the various judges, as is not based on any one particular theory of creativity (Amabile, 1982). The CAT does not directly address domain general versus domain specific creativity, but creative artefacts from a particular domain are judged by experts from that field. As Csikszentmihalyi (1999) discusses, creativity entails a process that produces a new idea or artefact that is recognized as being creative by others. The argument behind this approach is that the best reference for judging creativity comes from those with in-depth knowledge of the area. Because critics can have differing opinions, the assessments are pooled across judges. That said, the technique has been shown to demonstrate reasonable reliability and reproducibility across sessions (Baer & McKool, 2009). An adaptation of the CAT has successfully been used in the assessment of the creativity of computationally-generated music (Pearce & Wiggins, 2007). This approach is useful for eliminating judges' bias against artefacts generated from artificial systems. One limitation of the CAT, however, is that the process leading up to the creative artefact is not assessed.

Perspectives and experimental tasks from psychology

The previous section identified ways that creative thinking and production is assessed, with a focus placed on output; that is, examining generated solutions or a creative product. The following sections will address creativity in terms of both output and the (cognitive) processes underlying creative generation. Arguably, the "under the surface" mechanisms that give rise to creative behaviour are just as important to creativity as the output itself, *especially* in regard to computational creativity.

So then, which mechanisms are relevant? Arguably, learning and memory are fundamentally important for creativity. Whether humans intentionally generate new creative ideas or products out of existing knowledge, or have "eureka" moments of seemingly sudden insight, a detailed knowledge of the domain is required. The creator must first have learned the structure or statistics defining the domain before he may combine existing ideas in novel ways. Put a different way, the creator must first learn examples and rules defining the relevant conceptual spaces (see Gärdenfors, 2000; and Boden, 1990). Creative processes can then be thought of as finding new concepts, exploring previously unchartered areas of a conceptual space, or transforming the dimensions of the space. Because learning and manipulating concepts is essential to creativity, memory representations can be considered to be the building blocks of the creative process. Further, some focus will be placed on attention and expectation mechanisms, as these underlie human perception and memory.

The importance of anticipation, prediction, expectation in creativity and assessing creative output

Temporal processing of sequences, such as music and language, relies heavily upon prediction. Expectation is crucial for learning, and influences which information enters and updates memory; therefore, prediction mechanisms are integral to creativity. Our brain picks up patterns in our environment from which we make predictions about the world. For example, while listening to speech, one does not passively "hear" words go by; rather, the listener makes implicit predictions about what will happen next. For example, in the sentence, "I saw the bird flap its….", the listener will have a strong expectation for "wings." If an unexpected word occurs instead, such as "banana", a surprise response to the semantic incongruity is elicited in the brain (e.g., see Kutas & Hillyard, 1984). Through correct and incorrect expectations (that is, expectations that are either validated or violated by the ongoing stream of information), we learn the rules and structure of the signal or stimulus. We are then able to form increasingly robust prediction mechanisms for the temporal sequence, and the learning process continues dynamically in this cycle of prediction, feedback, and updating of an internal predictive model.

There is a reason our perception and cognition are largely dependent on prediction: expectation is evolutionarily important; we must be able to anticipate the world in order to survive and thrive in our environment. Consider driving in traffic, where correct prediction does have life or death consequences, or the motor planning involved when reaching to lift a glass of water, for example. As alluded to before, prediction is also essential for the temporal processing of sequences, such as language and music. Through the prediction-feedback loop described above, we are able to learn a statistical framework and network of associations (analogies, scripts, schemas, etc) that guides our perception and cognition.

Arguably, expectation and contextual association occurs at every level of abstraction in the brain, from low-level processing that can be described by Bayesian relationships, to high-level areas in the neocortex responsible for categorization and concept formation. This paper will cite methods for testing these different kinds of prediction, whether statistical, semantic, syntactic, physiological, conceptual, or symbolic.

Correct prediction is a valuable evolutionary trait, and because we cannot always rely on direct experience to form predictions (for example, when encountering an unfamiliar member of the large cat family, we should probably not approach the wild animal), humans have developed affective responses to the validation or denial of predictions. Through this emotional response to prediction, we develop preferences, ranging from the mundane to fear and pleasure. In relation to phenomena such as novelty seeking, creativity, and aesthetics, hedonic preferences may often be described in terms of an inverted-U relationship as described by Wundt and Berlyne (see Berlyne, 1957; 1970). The hedonic function is as follows: as intensity or complexity increases, preference or pleasure increases, until a (subjective) point at the top of the Wundt curve, after which increasing complexity results in diminishing preference. Essentially, stimuli that are very predictable are perceived as boring, and stimuli that are extremely complex or intense are viewed as inaccessible or overstimulating. Therefore, the center of the curve yields a “sweet spot” of optimal complexity, which can be a function of the complexity/intensity of the stimuli, and the observer’s experience. This theory has been applied in a range of domains, including music perception and creativity research (e.g., Steck & Machotka, 1975; Martindale, Moore, & Borkum, 1990; Saunders & Gero, 2001; North, & Hargreaves, 2008).

The role of expectation is not new to the field of creativity, as demonstrated by the work of Wiggins and colleagues (Wiggins, Pearce, & Müllensiefen, 2009; Pearce & Wiggins, 2007) and Maher (2010). Pearce and Wiggins have used measures of unexpectedness in implementations of computational systems and in behavioural tests of music perception (Pearce & Wiggins, 2006; Pearce, et al., 2010). Maher speaks of expectedness and surprise as an evaluation metric – like many other approaches, she identifies novelty and value as mandatory criteria for evaluating creativity, but in addition to these she adds *unexpectedness*. Maher argues that whereas novelty is a metric for identifying how different a particular artefact is from other examples in its class, unexpectedness is based on a set of temporal expectations; it quantifies not how different the artefact is, but how much it deviates from expectation (Maher, 2010).

To summarize the relevance of expectation mechanisms on computational creativity: prediction helps the brain learn and encode information about a domain, and violation of expectation is related to aesthetics and affect. Creativity often involves the discovery of novel solutions to a problem or task; it is the act of exploring (finding new regions or pathways) or extending a learned space of mental representations. The network of mental representations or conceptual spaces may be thought of as a complex prior distribution, with statistically-defined co-occurrence and correlational features. In this framework, exploratory creativity involves generating new ideas based on learned probability distributions or conceptual representations, and preference for generated creative artefacts (value judgements) generally falls within a preferred range of familiar and novel, or predictable and complex.

Perceptual, cognitive and computational methods

Correlational studies and rating scales

Correlational studies use a statistical analysis to assess the relationship between two variables. The relationship is measured in terms of direction (positive or negative correlation) and strength (e.g., a correlation coefficient between 0 and 1). Although correlational studies do not provide evidence for causality, they can be a valuable indicator, and inspire empirical studies that can test causal relationships.

A potentially useful technique within computational creativity would be to correlate properties of the creative system or the creative artefact with behavioural responses from observers. Within the domain of music for example, listeners' ratings may be collected to assess cognitive and affective states in response to melodies varying in style or familiarity. Several different types of response scales may be used. A Likert scale is a psychometric rating scale in which participants respond on a symmetric response range from disagreement to agreement. For example, a participant may be asked to respond on a scale from 'Strongly Agree' to 'Strongly Disagree' to capture how strongly they believe a particular statement is valid. Other rating scales partition the scale numerically, asking, for example, "How creative [enjoyable/imaginative/novel] was this poem on a scale from 1 to 7?" Scales of this type have been utilized to prompt listeners about preference, expectation, perceived complexity, aesthetics, functionality, novelty, etc. Although subjective and perhaps seemingly unreliable, when administered correctly, ratings scales can provide very robust and consistent measures of participants' judgements (e.g., Preston & Colman, 2000).

Some techniques in creativity research have been devised to overcome the individual subjectivity of observers' ratings. The Consensual Assessment Technique (Amabile, 1982), for example, employs the independent evaluation of creative artefacts by those with experience in the relevant domain. By measuring the degree to which a group of assessments agree, the issue of individual subjectivity is resolved. The technique also utilizes measures of reliability (consistency in agreement among raters) and validity to ensure that the findings are robust and replicable.

One consideration while collecting behavioural ratings, especially from non-experts, is that asking observers to rate their views can alter their views (Schwarz, 1999). Therefore, another approach to assessing creativity is through indirect measures of perception and cognition, such as reaction time measurements, which can skirt this issue of bias. Indirect measures are also useful in cases where expert evaluation is not feasible.

Reaction time

Rather than collect ratings that ask participants to make a direct judgement, reaction time (RT) may be used as an indirect measure of perceptual or cognitive processing. RT is the time between the presentation of a stimulus and a subsequent behavioural response, such as a button press. This method may be used for a variety of tasks; for example, RT data may be used to assess the relative expectedness of stimuli that vary in predictability or group membership, with the hypothesis that deviant stimuli will result in longer reaction times.

Physiological response and motion capture

Physiological measurements can be used to capture physical manifestations of psychological states. (Note that psychophysiological measures of brain activity are discussed separately in this paper.) Common measures are heart rate and Galvanic skin response, an electrodermal response indicative of a person's physiological and psychological state of arousal (a term used to describe overall activity or energy). These techniques have been used to assess perceived tension, stimulus intensity, or evoked emotion in a range of domains.

Motion capture is another approach to measuring real-time indicators of emotion, arousal, and embodied cognitive states. In this technique, motion sensors are placed on the participant's body to record movement in real-time while performing a task. This technique has been used in studies of music performance, for example, to correlate performers' movements (sometimes during specific sections of music) to perceived emotion in the audience (Livingstone, Thompson, & Russo, 2009; Friberg, 2004). If testing the creativity of generated music, one could ask professional musicians to perform human and computer-generated music, while capturing performers' movements (again, indicating affective state) during both types of performances. One could then examine whether movement differed between the types of generated music.

Theoretical models of memory

Because memory representations form the building blocks for combining existing knowledge into new ideas, psychological theories of memory are worth mentioning. There are many psychological models of memory; it is not the function of this paper to review all of these, but rather to give readers (especially those who are not familiar with this area) a few examples to provide a framework for understanding the usefulness of these paradigms, and applicability of these approaches to computational models of creativity.

The dual store memory model (Atkinson & Shiffrin, 1968) proposed that as a stream of incoming information is perceived, items are temporarily stored in short-term memory and associations are reinforced in long-term memory. As new information is presented, old items must move out of the short-term memory buffer to make room for new items. But each instance of recalling an item (rehearsal in short-term memory) creates stronger associations in long-term memory. This model may inform computational cognitive architectures by suggesting mechanisms and the dynamics of information flow.

In another influential model of short-term memory, Baddeley and Hitch (1974), an executive control system monitors and controls several component systems, including a phonological loop, visuo-spatial sketchpad, and episodic buffer that processes incoming information. The way that these components interact can inform multi-modal models of computationally creative agents.

<u>Conceptual spaces</u>

Conceptual spaces theory (Gärdenfors, 2004) offers a geometrical model of concept representation, in which points (objects) lie within geometrically defined regions (concepts). A conceptual space is comprised of a set of *quality dimensions*, such as color, pitch, or force, and the dimensions may be physical or abstract. These dimensions correspond to ways of comparing objects, namely, how similar (proximal) or distant stimuli are within a space. For example, our conceptual space for color is defined by the three dimensions of hue, saturation, and brightness, all of which can be described in terms of geometrical shapes. Certain dimensions are integral; this signifies that giving value to one dimension (eg, pitch) necessitates value in another dimension (loudness) – if a pitch exists, it must have a certain loudness. Other dimensions are separable, such as size and hue. Gärdenfors defines a *domain* as a collection of integral dimensions that are separable from every other dimension. This yields a slightly more specific definition of a conceptual space as a set of quality dimensions that are divided into domains. More formalizations are present to define the geometrical characteristics of domains and regions, such as *natural properties*, but for a more in-depth account I refer the reader to Gärdenfors (2004).

There are several ways one can test hypotheses derived from conceptual spaces theory. One testable hypothesis is that it should be cognitively easier (and less creative) to identify new points (ideas/examples) within a conceptual space rather than change the geometry of the space (as in transformation creativity). One way of experimentally testing the novelty of conceptual creation is to measure participants' reaction time while comparing, for example, points within one natural property (defined by a convex region of a domain), versus points spanning multiple domains.

A recent paper suggests the use of a formal implementation based on Gardenfors' conceptual spaces to draw analogies between different domains, namely musical pitch perception and visual scene perception (Chella, et al., 2014). Conceptual spaces theory requires validation as a cognitive model, not through the use of subjective measures, but through empirical testing.

Veale, Gervás, and Pérez y Pérez (2010) address exploratory behavior of conceptual spaces in the context of creative story generation. Because an artificial system may generate myriad story ideas, many will be of poor quality. Therefore, a key component of their approach is to reject candidate stories by confining the system's exploratory behavior to more highly-valued regions of spaces; in other words, exploration of conceptual space is constrained based on quality dimensions, such as emotion, interest, and tension (see León & Gervás, 2010).

<u>Information Theory</u>

Information theory measures the amount of information contained in a transmitted signal, and has been widely applied in fields as diverse as astronomy and linguistics. Measures such as entropy (amount of uncertainty) and information content (IC) (unexpectedness; the probability of an event given context) have been shown to reflect measures of perceptual and cognitive processing. For example, in music perception, Pearce and colleagues have shown that IC captures the unexpectedness of musical events (Pearce & Wiggins, 2006; Pearce, et al., 2010; Agres, Abdallah, & Pearce, 2013). By calculating the IC of discrete events, one can make predictions about which events should be the least or most expected (see Pearce, et al., 2010) in the sequence. For behavioural validation, participants can rate the expectedness of the events; indeed, this has been shown to be a robust measure for capturing participants' psychological expectation and surprise. Rather than solely examine point-wise measures of information, one can also test the effect of information properties of the entire sequence (see Agres, Abdallah, & Pearce, 2013). The information theoretic properties of whole sequences have been shown to have a dynamic effect on perception and memory over time, with complex stimuli often having an increasing effect of poor recognition memory performance.

Computational approaches

Machine learning approaches have an important role in modelling learning and creativity. From perceptrons, neural networks, and Markov models to more recent developments in deep learning such as autoencoders and stacked RBMs, each of these methods can offer a useful means of exploring the mechanisms underlying creativity. Deep belief networks have not explicitly been used to simulate creative behaviour, but as they seem very promising for this endeavour, emphasis will be placed upon discussing ways of using deep learning methods to model learning, expectation, creative generation, and evaluation of creative output.

Perceptrons (Rosenblatt, 1957) are a type of biologically-inspired linear classifier, one of the earliest methods for using a classification algorithm to map real-valued data to categorical (and in this case, binary) output. Perceptrons were a building block for more sophisticated methods, such as artificial neural networks (ANNs), that have played a crucial role in machine learning. ANN contain an input layer, one or more hidden and context layers, and an output layer. The hidden layers allow the network to learn patterns in the input, extract patterns, and transform the information (make predictions or classify the input) into meaningful output. This method is clearly applicable to the study of creativity, both in learning representations, and in manipulating those representations to form novel output patterns. Some ANNs are feed-forward, and some utilize backpropogation of errors as a type of supervised learning internal feedback (comparing the network's predictions about the next event with the actual next event). ANNs may also utilize a recurrent "context" layer that serves as a history or memory for previously seen events. With the greater processing power of modern computing, and updated algorithms and other techniques (see for example the Dropout method proposed by Hinton, et al., 2012), these types of networks have become a popular research focus again, and their development has contributed greatly to advances in deep learning methods.

Deep learning methods

Deep learning is a machine learning approach to learning distributed representations that can yield high-level features or abstractions of data. Deep models contain multiple layers in-between the input and output layers (or involve stacking visible and hidden layers), and through the use of unsupervised learning and non-linear transformations, each successive layer reduces the dimensionality of the previous layer, which then reflects a higher level of abstraction. Various types of deep architectures exist, including deep neural networks, deep belief networks, convolutional neural networks, and recurrent neural network-restricted Boltzman machines (RNN-RBMs), to name a few.

There is evidence that this hierarchical progression of abstract representation mimics that found in the brain (e.g., Van Essen & Maunsell, 1983). Also, during human development there is a trend of moving from literal interpretations to increasingly abstract conceptualizations. For example, as noted by Healey et al. (2010), children's drawings typically progress from scribbling, figurative drawings (i.e., of people and animals), and concrete representations, to more economic, meaningful, and symbolic representational drawings (Karmiloff-Smith, 1979). This developmental change reflects the acquisition of more abstract cognitive processes. Therefore, deep learning models may offer new ways of describing, exploring, and forming predictions about the acquisition of increasingly abstract mental operations, both in terms of development, and also in terms of the organization of information in the adult brain.

Because deep belief networks are capable of learning feature hierarchies in data, they offer a promising approach to computational creativity. As discussed previously, learning is a crucial prerequisite to creative behaviour in humans. Learning representations may also allow for more flexibility (different distributed representations may be combined at different levels of a model) and for more truly creative behaviour in artificial systems. For example, in the highest, most abstracted representations of the data, the relationship between abstracted features and associations may be explored. When projected into feature space, different features may be combined in novel ways, or the feature space itself may be explored and transformed to yield new creative objects/features/concepts.

One criticism of deep learning approaches has been the lack of a representation of time. Most deep models learn the conditional probability distributions within the data, but do not encode any temporal dynamics of the input. For data that are sequential in nature, such as music and language, this poses a significant constraint. Researchers have recently begun to explore new deep architectures that can both discover conditional probabilities and temporal patterns in the data (see Längkvist, Karlsson, & Loutfi, 2014), such as temporal RBMs (Sutskever & Hinton, 2006; Bengio, 2009), variations of convolutional RBMs (Bo, Ting, & de Freitas, 2010; LeCun & Bengio, 1995), multiplicative RNNs (Sutskever, Martens, & Hinton, 2011), and RNN-RBMs (Boulanger-Lewandowski, Bengio, & Vincent, 2012).

Several variables determine the success of deep models in abstracting meaningful patterns in data. These models are constrained by what they are able to represent, and therefore choosing the input representation has massive implications for the types of statistical relationships the network is able to learn. Also, the architecture of deep networks utilizing stacked RBMs is such that the low-level statistical information that is gleaned from the input in the first hidden layer is then used as input for the next layer in the network (Bengio, 2009). This next hidden layer can then learn statistical patterns from the statistical relationships from the prior layer, and thus learn correlational relationships at a higher, more abstract level. Different layers can learn different "features", which are based on reducing the number of dimensions that are used to describe the information.

Because this flow of dimensional reduction in representation happens in a step-by-step fashion (that is, one layer at a time), the features that a layer can learn depend entirely upon the statistics of the previous layer. This constraint limits the type of features that the system can learn. As an alternative, the information in a given layer could potentially be partitioned and then fed to multiple higher sub-layers, which would then find the statistical relationships in that set or partition of data. These multiple sub-layers can then be weighted and combined to discover even more abstract features. This could then give rise to combinatorial creativity.

Hierarchical Temporal Memory Model

Another approach that is complementary to deep neural networks is one spearheaded by Jeff Hawkins, using his concepts of Hierarchical Temporal Memory and a Cortical Learning Algorithm (Hawkins & Blakeslee, 2004). In this model, sparse, distributed representations encode symbolic information about the world that capture the temporal transitions and high level patterns between events. The system is biologically inspired, with multiple layers (reflecting the cortical layers of the neocortex), and columns through these layers that represent different symbolic elements of an entity (for example, the different characteristics of "dog", such as number of legs, sounds produced, what it eats, etc). Because there are sparse distributed representations, when a new example (say a zebra) is shown to the system, although the system will never have seen this example before, some of the characteristics of zebra will overlap with the characteristics of other animals. Therefore, some of the “on” bits that encode for those characteristics will be the same as other animals, and then the system can infer information about the zebra. In this way, one can imagine applications to creativity: rather than feed the model an animal whose sparse distributed representation is new, one can prompt the model to discover new sparse distributed networks that correspond to new animals (or more broadly, new sets of features that represent a new concept). This is also reminiscent of exploring a Gardenforsian conceptual space to find new points within that space (Gardenfors, 2004).

The Neuroscience of Creativity

Neuroscience methods such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) have tried to assess differences in brain activity while humans exhibit creativity (vs perform a non-creative task), and have also tested whether the brain activity of highly creative individuals is markedly different from that of relatively non-creative individuals. An in-depth review of EEG and imaging findings will not be provided here, as recent work has already achieved this goal (for reviews, see Dietrich & Kanso, 2010; and Jung, et al., 2010). An overview of approaches and findings will be provided, as well as a brief summary of works not included in the aforementioned reviews.

Electroencephalography

Electroencephalography (EEG) measures the electrical activity in the brain, as measured on the scalp. Populations of neurons, when firing together, emit enough electrical activity to be measured (albeit amplified) through the layers of meninges and cranium. It should be noted that EEG measures cortical activity; electrical signals from inner brain structures (such as the basal ganglia) dissipate too much by the time they reach the skull. Changes in electrical activity are thought to reflect underlying post-synaptic neural processing, with particular kinds of EEG (oscillatory activity, in different parts of the brain, mapping on to various perceptual or cognitive processing states. EEG methods are often not used for localization of function, as the spreading signal is broadened and muffled by the time it reaches the electrodes on the scalp. Rather, EEG has excellent temporal resolution, with measurements (on as many as 256 electrodes placed around the scalp) recorded as often as every 1 or 2 milliseconds.

Event related potential technique

One of the most common EEG techniques is event-related potential (ERP) analysis, which measures electrical activity (in terms of amplitude in μV and latency in ms) immediately following an event (i.e., an experimental stimulus) (see Fig.1).

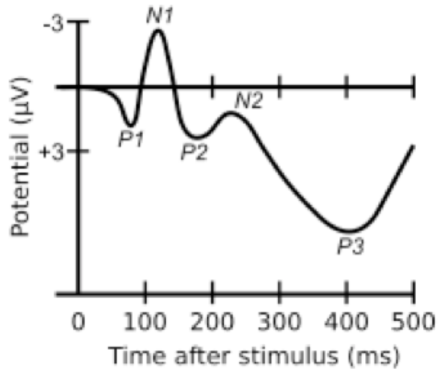

Fig.1 – Event Related Potential, with positive μV plotted downwards as per convention (image permission sought).

There are characteristic waves in the ERP response: First the N1 is a large negative-going evoked potential, which peaks in amplitude around 80-120vms post-stimulus onset. This activation is usually distributed in frontal-central areas. This wave is followed by a positive-going evoked potential (the P2), which peaks around 150-270 ms post-stimulus onset. Researchers make a distinction between the auditory evoked response and visual responses, for example, which display small variations in amplitude and latency. The N1 is a preattentive response that is sensitive to the unexpectedness of an event (Lange, 2009), as well as physical properties of the stimulus (in the case of audition, the loudness or frequency of the stimulus, for example). In addition to stimulus predictability, the amplitude of the N1 can also reflect focused attention (Luck, Heinze, Mangun, & Hillyard, 1990; Alho, et al., 1997). Due to the noisy nature of EEG data and data collection, many trials (depending on the paradigm, typically a minimum of 80 trials per individual) are needed to compute an average ERP to a particular type of stimulus (e.g., a deviant tone).

In addition to the N1-P2 complex, certain other ERP methods look at different kinds of evoked responses. In linguistic tasks, the P6 is elicited in response to a syntactic anomaly (Osterhout & Holcomb, 1992), and the N4 occurs when there is a semantic incongruity (Kutas & Hillyard, 1980). The mismatch negativity (MMN) is a difference wave occurring around 150-250 ms post-stimulus onset to a deviant event; for example, in a series of repeated tones, an oddball (unexpected or deviant) tone will elicit an MMN response, regardless of whether the listener is attending the stimulus stream (see Näätänen, Paavilainen, Rinne, & Alho, 2007, for a review). The auditory MMN can be evoked from a change in pitch, stimulus location, loudness, duration, or spectral content (Friedman, Cycowicz, & Gaeta, 2001). The P3 component (250-450 ms post stimulus onset) is implicated in decision making or stimulus evaluation, and is also sensitive to unexpected or deviant stimuli (Alho, et al., 1997; Polich, 1986).

In terms of creativity research, ERP components can be used to assess expectation mechanisms and semantics. As discussed above, the amplitude and latency of the N1 component is modulated by unexpected events occurring within a predictable context. Because novelty is an important aspect of evaluating creativity, a neural response to unexpected stimuli is a useful tool for assessing perceived surprise.

Evidence also exists that N4 is sensitive to predictability: the amplitude of these components may increase during rule-learning, as when participants are implicitly learning the transitional probabilities between tones within a "tritone word" (a set of three tones, analogous to a word with three phonemes) (Abla, Katahira, & Okanoya, 2008). The N4, then, may be used as a proxy for statistical learning or segmentation. As discussed above, segmentation ability reflects the degree of learning; therefore, this measure can be used, for example, to compare computational segmentation to neural signatures of segmentation in humans.

EEG time-frequency analysis

Researchers also commonly examine temporal and spatial activation over broad swaths of the cortex. In time-frequency analyses, global oscillatory activity is assessed in terms of different bands of activity. The theta band (4-7Hz) is seen during drowsiness and states of low physiological arousal, but has also been connected to suppression and inhibition, which is exhibited, for example, when an individual is trying to focus on an auditory stimulus and ignore a visual stimulus. Alpha activity (8-12 Hz) has been widely studied, with some controversy of findings. It is commonly associated with calm states; posterior alpha waves (distributed over the occipital lobe) while the individual's eyes are closed is often indicative of relaxation or drowsiness. Like theta activity, some research argues that more alpha power reflects passive or active inhibition of non-relevant information while performing a task. Beta band activity (13-30 Hz) typically reflects active, alert states.

One approach in the neuroscience of creativity has been to examine oscillatory band activity during creative compared with non-creative tasks. One hypothesis is that alpha activity supports divergent thinking (Martindale 1990), and therefore more alpha and less beta band power should be present while individuals perform a creative task. Although this hypothesis has found some support, other studies have not seen this elevation in alpha power for creative versus non-creative activities (see Dietrich & Kanso, 2010). Future research in this direction should carefully break down creativity into more specific testable hypotheses.

Transcranial Magnetic Stimulation

Transcranial Magnetic Stimulation, or TMS, uses an electromagnet placed over the scalp to apply electromagnetic stimulation to a localized part of the brain. The method focally alters the activation of neurons in the underlying brain area, which serves to temporarily inhibit normal functionality. Because the effect is reversible, TMS can be used to examine localization of function, and is often applied to motor cortex or frontal areas. A similar method, transcranial direct current stimulation (tDCS), uses two electrodes placed on the scalp to modulate brain activity, based on the location and polarity of the electrodes.

TMS has been applied to frontal areas to inhibit frontal areas; limiting frontal top-down control can have the effect of enabling artistic, creative, or even savant-like abilities (eg., Snyder et al, 2003; Snyder, et al., 2006). Facilitation of expert-like generation through frontal inhibition is based around the somewhat counterintuitive hypothesis that latent savant-like abilities are present within us all, but are not normally demonstrated due to inhibitive top-down processing constraints. Obstructing this inhibitive information from high-level areas (semantic and abstracted, or executive control areas) may be thought of as constraining convergent thinking. Put differently, repetitive Transcranial Magnetic Stimulation (rTMS) may provide a means of facilitating divergent thinking ability.

In addition to temporarily enhancing creative ability, rTMS has been used to identify brain areas/cognitive mechanisms underlying certain types of creativity. For example, when rTMS is applied to the right posterior superior temporal sulcus, understanding of novel metaphors is inhibited (as opposed to conventional metaphors or non-metaphorical word pairs) (Pobric, et al., 2008). rTMS applied to the left hemisphere impaired processing of conventional and literal words pairs, providing a functional double-dissociation between location (left or right hemisphere) and type of metaphorical processing (literal/conventional and novel metaphorical).

Other TMS work has localized functional brain regions that are involved with language versus musical processing. For example, TMS applied to left frontal areas results in impairment on speech production tasks for most people (Stewart, Walsh, Frith, & Rothwell, 2001). (It should be noted that the researchers confirm that this is not due to motor interference.) Interestingly, when TMS is applied to areas that result in speech impairment, participants remain able to sing melodies, without any impairment to the lyrics. Even when TMS is administered over the homologous area in the right hemisphere, song production remained intact in the test subjects. This may be because the neural mechanisms responsible for melodic production are more widely distributed than speech circuitry, or because speech mechanisms are more susceptible to interference than melodic production (Stewart, et al., 2001). In other words, we may be able to use this technique to isolate specific components and mechanisms of creative production in humans. We then have the opportunity to use these components to inform computational models of creativity, at least at a conceptual level. In addition, the above research speaks to creativity in that musical content allowed for intact production of words (lyrics) while singing – a different domain. In other words, information from one domain or conceptual space can not only facilitate creative behaviour in another domain, but make generation feasible where it was previously impossible. This is in contrast to the previous findings that *inhibition* of frontal areas can allow for more expert performance. Therefore, depending on the type of creativity, inhibition of particular mechanisms (e.g., high-level semantic knowledge) or facilitation through the use of additional information/mechanisms, can yield more creative behaviour.

Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is a technique that is used to examine brain activity by measuring changes in cerebral blood flow. The basic assumption is that neural activity is coupled with hemodynamic response, and that when a particular area of the brain is in use, oxygenated blood flow will increase in that region. Although fMRI has excellent spatial resolution (on the order of millimeters), the temporal resolution is poor (on the order of seconds), due to the time required by the hemodynamic response itself.

Imaging techniques been used to locate the areas and pathways implicated in various mental processes, from the phonological processing of speech (Price, 2012) to the neural correlates of emotion (Phan, Wager, Taylor, & Liberzon, 2002). To examine creativity in humans, fMRI has been used to identify the brain areas underlying creative processes such as metaphorical reasoning (Mashal, Faust, Hendler, & Jung-Beeman, 2007), jazz improvisation (Limb & Braun, 2008), and divergent thinking and creative story generation (Howard-Jones, Blakemore, Samuel, Summers, & Claxton, 2005). Taken as a whole, imaging studies have not verified the hypothesis that the right hemisphere is more reliably implicated in creative processing than the left hemisphere, or that certain brain regions are associated with creativity in general, with the exception of the prefrontal cortex (Dietrich & Kanso, 2010). Prefrontal activation is often mediated by creative compared with non-creative tasks, with the type of activation a function of the particular paradigm (e.g., Divergent thinking tests or jazz improvisation) (Fink, et al., 2009; Limb & Braun, 2008). In sum, fMRI findings suggest that creativity is not a stand-alone phenomenon that can be attributed to particular brain areas or pathways, but that particular creative behaviors may show reliable neural signatures (Dietrich & Kanso, 2010).

Conclusion

This paper provides an overview of behavioural, cognitive, computational, and neuroscientific methods that may be employed or adapted for the study of computational creativity. Alternative definitions of creativity and computational creativity were offered as a means of isolating particular aspects of creativity that may be systematically evaluated. In addition, theoretical arguments were given concerning the importance and relevance of learning, expectation, and memory to creativity. In essence, prior to a creative undertaking, pertinent information must be learned and represented. Methods ranging from deep learning to conceptual spaces theory to fMRI were cited as a means of testing the nature of these learned representations. The diversity of methods may be used to empirically investigate how existing knowledge may be combined in novel ways, how new ideas may be discovered, or how the trajectory through a creative mental space may change as a function of experience or motivation. It is our belief that different approaches and scientific tools may provide converging methods for a richer understanding of creativity. We hope this paper will promote more widespread use of empirical testing of creative processes and artefacts to support the development of computationally creative systems and theories of creativity.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20,* 952-964.

Agres, K., Abdallah, S., & Pearce, M. (2013). An Information-Theoretic Account of Musical Expectation and Memory. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 35th Annual Conference of the Cognitive Science Society (pp. 127-132). Austin, Texas: Cognitive Science Society.

Alho, K., Excera, C., Diaz, R., Yago, E., & Serra, J. (1997). Effects of involuntary auditory attention on visual task performance and brain activity. *NeuroReport, 8,* 3233-3237.

Amabile, T. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43,* 997-1013.

Amabile, T. (1983). *The social psychology of creativity.* New York: Springer-Verlag.

Amabile, T. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology, 48,* 393-399.

Amabile, T. (1996). *Creativity in Context*. Boulder, Colorado: Westview Press.

Baer, J., & McKool, S. (2009). Assessing creativity using the consensual assessment technique. In C. Schreiner (Ed), *Handbook of assessment technologies, methods, and applications in higher education.* Hershey, Pennsylvania: IGI Global.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning, 2,* 1-127.

Berlyne, D. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics, 8*, 279-286.

Bo, C., Ting, J., & de Freitas, N. (2010). Deep learning of invariant spatio-temporal features from video. In *NIPS Deep Learning and Unsupervised Feature Learning Workshop*, 2010.

Boden, M. (1990). *The creative mind: Myths and mechanisms.* New York: Routledge.

Boulanger-Lewandowski, N., Bengio, Y., Vincent, P. (2012). Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription. In *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK.

Chella, A., Gaglio, S., Oliveri, G., Augello, A., & Pilato, G. (2014). Creativity in conceptual spaces. In *Proceedings of the 5th International Conference on Computational Creativity*. Ljubljana, Slovenia. June 2014.

Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R.J. Sternberg (Ed.) *Handbook of Creativity*, Cambridge University Press, Cambridge, UK, 313-335.

Colton, S. (2008). Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI*, Spring Symposium on Creative Systems, 2008.

Colton, S., Charnley, J., & Pease, A. (2011). Computational creativity theory: The FACE and IDEA descriptive models. In *Proceedings of the 2nd International Conference on Computational Creativity*.

Colton, S., Pease, A., Corneli, J., Cook, M., & Llano, T. (2014). Assessing progress in building autonomously creative systems. In D. Ventura (Ed), *Proceedings of the Fifth International Conference on Computational Creativity*.

Dietrich, A., & Kanso, R. (2010). A Review of EEG, ERP, and Neuroimaging Studies of Creativity and Insight. *Psychological Bulletin, 136*, 822–848.

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities.* New York: Basic Books.

Fink, A., Grabner, R., Benedek, M., Reishofer, G., Hauswirth, V., Fally, M., Neuper, C., Ebner, F., & Neubauer, A. (2009). *Human Brain Mapping, 30,* 734-748.

Friberg, A. (2004). A fuzzy analyzer of emotional expression in music performance and body motion. In J. Sundberg & B. Brunson (Eds.) *Proceedings of Music and Music Science*, Stockholm.

Friedman, D., Cycowicz, Y., & Gaeta, H. (2001). The novelty P3: an event-related potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews, 25,* 355-373.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought.* Boston, MA: MIT Press.

Gorder, W. (1980). Divergent production abilities as constructs of musical creativity. *Journal of Research in Music Education, 28,* 34-42.

Guilford, J. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Hawkins, J., & Blakeslee, S. (2004). *On Intelligence: How a New Understanding of the Brain will Lead to the Creation of Truly Intelligent Machines.* New York: Times Books.

Healey, P., Swoboda, N., Umata, I., & King, J. (2010). Graphical language games: Interactional constraints on representational form. *Cognitive Science, 31,* 285-309.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *Technical report*, arXiv:1207.0580.

Howard-Jones, P., Blakemore, S.J., Samuel, E., Summers, I., & Claxton, G. (2005). Semantic divergence and creative story generation: An fMRI investigation. *Cognitive Brain Research, 25*, 240-250.

Jordanous, A. (2012). *Evaluating computational creativity: A standardised procedure for evaluating creative systems and its application*. Ph.D. Dissertation, University of Susex.

Karmiloff-Smith, A. (1979). Micro- and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science, 3*, 91-118.

Khatena, J., & Torrance, E. (1976). *Manual for Khatena-Tarrance Creative Perception Inventory*. Chicago: Stoelting.

Kim, K. (2006). *Can We Trust Creativity Tests? A Review of the Torrance Tests of Creative Thinking (TTCT)*. Creativity Research Journal, 18, 3–14.

Koestler, A. (1964). *The act of creation*. United Kingdom: Hutchinson.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161-163.

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition*, *69*, 127-137.

Längkvist, M., Karlsson, L., & Loutfi. A. (2014). A Review of Unsupervised Feature Learning and Deep Learning for Time-Series Modeling." Pattern Recognition Letters.

LeCun, Y., & Bengio, Y. (1995). *Convolutional networks for images, speech, and time series*. The handbook of brain theory and neural networks, 3361.

León, C., & Gervás, P. (2010). The role of evaluation-driven rejection in the successful exploration of a conceptual space of stories. *Minds & Machines, 20*, 615-634.

Limb, C., & Braun, A. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE, 3,* e1679. doi:10.1371/journal.pone.0001679

Livingstone, S., Thompson, W., & Russo, F. (2009). Facial expressions and emotional singing: A study of perception and production with motion capture and electromyography. *Music Perception, 26,* 475-488.

Luck, S., Heinze, H., Mangun, G., & Hillyard, S. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalography and clinical neurophysiology, 75,* 528-542.

Maher, M. (2010). Evaluating creativity in humans, computers, and collectively intelligent systems. In *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, 22-28.

Martindale, C., Moore, K., & Borkum, J. (1990). Aesthetic preference: Anomalous findings for Berlyne's psychobiological theory. *American Journal of Psychology, 103*, 53-80.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain and Language, 100,* 115-126.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220-232.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology 118,* 2544-2590.

North, A., & Hargreaves, D. (2008). *The social and applied psychology of music.* Oxford University Press, Oxford.

Osterhout, L., & Holcomb, P. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785-806.

Pearce, M., Ruiz, M., Kapasi, S., Wiggins, G., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage, 50,* 302-313.

Pearce, M., & Wiggins, G. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23*, 377-405.

Pearce, M., & Wiggins, G. (2007). Evaluating cognitive models of musical composition. In A. Cardoso and G. Wiggins (eds.) *Proceedings of the 4th International Joint Workshop on Computational Creativity*, 73-80.

Pobric, G., Mashal, N., Faust, M., & Lavidor, M. (2008). The role of the right cerebral hemishphere in processing novel metaphoric expressions: A transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience, 20,* 170-181.

Polich, J. (1986). Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 63,* 251-259.

Preston, C., & Colman, A. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104,* 1-15.

Price, C. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage, 62,* 816-147.

Phan, K., Wager, T., Taylor, S., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage, 16,* 331-348.

Ritchie, G. (2001). Assessing Creativity. In *Proceedings of the AISB Symposium on AI and Creativity in Arts and Science*, York, 2001.

Ritchie, G. (2007). Some Empirical Criteria for Attributing Creativity to a Computer Program. *Minds & Machines, 17*, 67–99.

Saunders, R., & Gero, J. (2001). Artificial creativity: A senthetic approach to the study of creative behaviour. In Gero, J.S. (ed.) *Proceedings of the Fifth Conference on Computational and Cognitive Models of Creative Design*. Key Centre of Design Computing and Cognition, Sydney.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93-105.

Shepard, R. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function, I. *Psychometrika, 27*, 125–140.

Snyder, A., Bahramali, H., Hawker, T., & Mitchell, J. (2006). Savant-like numerosity skills revealed in normal people by magnetic pulses. *Perception, 35,* 837–845.

Snyder, A., Mulcahy, E., Taylor, J., Mitchell, D., Sachdev, P., & Gandevia, S. (2006). Savant-like skills exposed in normal people by suppressing the left fronto-temporal lobe. *Journal of Integrative Neuroscience, 2,* 149-158.

Steck, L., & Machotka, P. (1975). Preference for musical complexity: Effects of context. *Journal of experimental psychology: Human perception and performance, 104*, 170-174.

Stewart, L., Walsh, V., Frith, U., & Rothwell, J. (2001). Transcranial magnetic stimulation produces speech arrest but not song arrest. *Annals of New York Academy of Sciences*, 433-435.

Sutskever, I., & Hinton, G. (2006). Learning multilevel distributed representations for high-dimensional sequences. *Technical Report.* University of Toronto.

Sutskever, I., Martens, J., & Hinton, G. (2011). Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning*, WA, USA.

Torrance, E. (1990). *The Torrance tests of creative thinking norms—technical manual figural (streamlined) forms A & B*. Bensenville, IL: Scholastic Testing Service, Inc.

Turner, M., & Fauconnier, G. (1995). Conceptual integration and formal expression. *Metaphor and Symbolic Activity, 10*, 183-204.

Van Essen, D., & Maunsell, J. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences, 6,* 370-375.

Veale, T., Gervás, P., and Pérez y Pérez, R. (2010). Computational creativity: A continuing journey. *Minds & Machines, 20*, 483-487.

Vold, J. (1986). *A study of musical problem solving behavior in kindergarten children and a comparison with other aspects of creative behaviour*. Doctoral dissertation, University of Alabama, Tuscaloosa.

Wallach, M., & Kogan, N. (1965). *Modes of thinking in young children*. New York: Holt, Rinehart and Winston.

Wallace, G. (1926). *The art of thought.* New York: Harcourt Brace.

Webster, P. (1987). Conceptual bases for creative thinking in music. In J. Craig Peery et al. (eds.), *Music and Child Development*, 158-174. New York: Springer-Verlag.

Wiggins, G. A. (2006a). A Preliminary framework for description, analysis and comparison of creative systems. *Journal of Knowledge-Based Systems, 19(7),* 449-458.

Wiggins, G. A. (2006b). Searching for Computational Creativity. *New Generation Computing, 24(3), 209-222*.

Wiggins, G., Pearce, M., & Müllensiefen, D. (2009). Computational modelling of music cognition and musical creativity. In, R. Dean (Ed.) *Oxford Handbook of Computer Music and Digital Sound Culture*. Oxford, UK: Oxford University Press.