Modelling creativity: identifying key components through a corpus based approach

Anna Jordanous[1,☯,*], Bill Keller[2,☯]

1 School of Computing, University of Kent, Chatham Maritime, Kent, UK
2 Department of Informatics, University of Sussex, Falmer, Brighton, UK

☯These authors contributed equally to this work.
*** a.k.jordanous@kent.ac.uk / billk@sussex.ac.uk**

Abstract

~~Creativity is a complex multi-dimensional concept which resists definition; it is accessible to us at a subjective, multi-faceted and dynamic level rather than being fixed and clearly defined. If we wish to study creativity scientifically using computational models, or simulate it computationally, then a tractable and accurate interpretation of creativity is needed. Creativity encompasses many related aspects, abilities, properties and behaviours. This paper uses techniques from the field of statistical natural language processing to identify a collection of fourteen key components of creativity through an analysis of what is considered to be important in talking about creativity. Words were identified which appeared significantly often in connection with discussions of the concept. Using a measure of lexical similarity to help cluster these words, a number of distinct themes emerged, which collectively contribute to a comprehensive and multi-perspective definition of creativity. The components provide an ontology of creativity: a set of building blocks which, in evaluation, have proven useful for making the concept easier to understand and more tractable to study and evaluate.~~

Creativity is a complex, multi-faceted concept encompassing a variety of related aspects, abilities, properties and behaviours. If we wish to study creativity scientifically, then a tractable and well-articulated model of creativity is required. Such a model would be of great value to researchers investigating the nature of creativity and in particular, those concerned with the evaluation of creative practice. This paper describes a unique approach to developing a suitable model of how creative behaviour emerges that is based on the words people used to describe the concept. Using techniques from the field of statistical natural language processing, we identify a collection of fourteen key components of creativity through an analysis of a corpus of academic papers on the topic. Words are identified which appear significantly often in connection with discussions of the concept. Using a measure of lexical similarity to help cluster these words, a number of distinct themes emerge, which collectively contribute to a comprehensive and multi-perspective model of creativity. The components provide an ontology of creativity: a set of building blocks which can be used to model creative practice in a variety of domains. The components have been employed in two case studies to evaluate the creativity of computational systems and have proven useful in articulating achievements of this work and directions for further research.

Introduction

What is creativity, and how can we better understand and learn about creativity using computational modelling? Computational creativity is a relatively youthful research area that has been growing with significant pace in recent years. 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.' [1, p. 21]

Computational creativity research follows both theoretical and practical directions and crosses several disciplinary boundaries between the arts, sciences and engineering. Research within the field is influenced by artificial intelligence, computer science, psychology and specific creative domains that have received attention from computational creativity researchers to date, such as art, music, reasoning and narrative/story telling [2–5, provide examples].

The evaluation of creative systems developed by researchers in the field of computational creativity has proven non-trivial. Creativity evaluation, a recurring topic for discussion, has been described as a 'Grand Challenge' for computational creativity research [6]. Difficulties are inherently linked to a question that both motivates and complicates the computational modelling of creativity: what do we mean when we talk about 'creativity' and what does it constitute?

Creativity is a complex, multi-faceted concept encompassing a variety of related aspects, abilities, properties and behaviours. There have been many attempts to capture this concept in words; indeed the work described in this paper is based on thirty such attempts (see the Methods section and the papers listed in Fig. 1). In the academic literature on creativity, many common themes have emerged. However, multiple viewpoints exist, prioritising different aspects of the concept according to what are traditionally considered to be the primary factors for a particular discipline. The need for a more over-arching, inclusive, multi-dimensional account of creativity has been widely recognised [7–10]. Such a meta-level account would assist our understanding of creativity, highlighting areas of common ground and avoiding the pitfalls of disciplinary bias [11, 12].

There are many challenges to modelling a concept like creativity in a computational setting. Conceptually, creativity seems inherently fuzzy or vague, with a meaning that shifts depending on the domain of application. Tackling these challenges affords two key advantages, both of which motivate the current paper. First, we can take advantage of computing and artificial intelligence to perform and/or enhance creative activities using computational power and research expertise. Secondly, the act of modelling creativity requires us to more carefully identify what informs our intuitive notions about creativity and this can guide us towards a more rigorous and comprehensive understanding of the concept.

The aim of the work reported in this paper is to examine the nature of creativity and to identify within it a set of components, representing key dimensions, that are recognised across a combination of different viewpoints. We present a novel, empirical approach to the problem of modelling how creative behaviour emerges, that focuses on what is revealed about our understanding of creativity and its attributes by the words we use to discuss and debate the nature of the concept. Analysis of this language provides a sound basis for constructing a sufficiently detailed and comprehensive model of creativity [13, 14]. The current work is intended as a significant, methodological contribution towards addressing the Grand Challenge of evaluation in computational creativity research. It should provide researchers with a firm foundation for evaluating exactly how creative so-called *c*reative systems actually are.

On our approach, statistical language processing techniques are used to identify words significantly associated with creativity in a corpus of academic papers on the subject. A corpus spanning some 60 years of research into the nature of creativity was collected together. The papers were gathered from a wide variety of disciplines including psychology, educational testing and computational creativity, amongst others. The language data drawn from this collection was then analysed and contrasted with data from a corpus of matched papers on subjects unrelated to creativity. From this analysis, a set of 694 *creativity words* was identified, where each creativity word appeared significantly more often than expected in the corpus of creativity papers. A measure of lexical similarity provided a basis for clustering the creativity words into groups of words with similar or shared aspects of meaning. Through inspection of these clusters, a total of fourteen *key components of creativity* was identified, where each represents a key theme or attribute of creativity. The set of components yields information about the nature of creativity, based on what is collectively emphasised in discussions about the concept.

In the rest of this section we begin by noting a variety of attempts to define creativity. The representation of subjective, ambiguous, loosely structured concepts is considered. In the remaining sections, details are provided of the methodology used to identify components of creativity from an analysis of language data. The results of this analysis are then presented in terms of a model that encompasses fourteen key components. The derived set of components is evaluated in terms of how well it satisfies the need for a shared, inclusive and comprehensive account of creativity and provides a vocabulary of creativity that is accessible to both people and machines. Finally, conclusions are drawn and some directions for further work are outlined.

Background: *The nature of creativity*

As Torrance observes:

> '[c]reativity defies precise definition ... even if we had a precise conception of creativity, I am certain we would have difficulty putting it into words' [15, p. 43].

Many other authors have expressed similar difficulties [7, 10, 16].

~~**A question of definition?** Plucker makes the case that the lack of a common understanding for creativity research weakens the 'legitimacy' and validity of that research. He notes that 'Without an agreed-on definition of the construct, creativity's potential contributions to psychology and education will remain limited' [9, p.87] and further that 'unless the definitional problem is addressed, creativity research will continue to be impeded by lack of direction, damaging mythologies, and general misunderstanding' [9, p.92]. In fact, 'Rather than being a strength of the field, as many believe, the lack of a common definition is a major, debilitating weakness. ... we feel that an agreed-on definition is long overdue and has placed the field in a crisis of legitimacy. ... This change in the focus and direction of creativity research is needed if the field is to move from a shadowy past into the forefront of constructive approaches' [9, p. 93]. Other researchers share these concerns: 'I submit that the time has come for more precision in definition and usage [of the word creativity], that only when the field is analyzed and organized - when the listener can be sure he knows what the speaker is talking about - will the pseudo aspect of the subject of creativity disappear' [7, p. 310].~~

In their review of research into human creativity, Hennessey and Amabile ask a significant follow-on question:

> 'Even if this mysterious phenomenon can be isolated, quantified, and dissected, why bother? Wouldn't it make more sense to revel in the mystery and wonder of it all?' [11, p. 570]

Two answers to this question are offered by Hennessey and Amabile, both of which are identified as desirable: to gain a deeper understanding of creativity and to learn how to boost people's creativity.

Creativity can and should be studied and measured scientifically, but the lack of a commonly-agreed understanding causes problems for measurement [10]. Plucker et al. make recommendations about best practice based on their own survey of the creativity literature:

> 'we argue that creativity researchers must
>
> (a) explicitly define what they mean by creativity,
>
> (b) avoid using scores of creativity measures as the sole definition of creativity (e.g., creativity is what creativity tests measure and creativity tests measure creativity, therefore we will use a score on a creativity test as our outcome variable),
>
> (c) discuss how the definition they are using is similar to or different from other definitions, and
>
> (d) address the question of creativity for whom and in what context.' [9, p.92]

In short, we need to specify and justify the standards that we use to judge creativity. A more objective and well-articulated account of how creativity is manifested enables researchers to make a worthwhile contribution [8–10]. **Particularly, in research we would like to focus on what processes and concepts relevant to creativity are 'sufficiently important to warrant study' [17, p. 15], based on an *accumulation* of the body of work on creativity to date [17].**

Definitions of creativity. To find out the meaning of a word, a natural first step might be to consult a dictionary. Dictionary definitions of 'creativity' provide a brief introduction to the meaning of the word. However, for the purposes of research, the utility of such definitions is severely restricted by their format and brevity, and they generally provide only cursory, shallow insights into the nature of creativity. More problematic still, dictionary entries are often self-referential or circular, defining creativity in terms of "being creative" or "creative ability". To illustrate these limitations, there follow several typical dictionary definitions of creativity and the related words creative and create:[1]

Oxford English Dictionary 2nd ed. (1989) pp. 1134-5:

creativity: creative power or faculty; ability to create

creative: Having the quality of creating, able to create; of or relating to creation; originative. b. Inventive, imaginative; of, relating to, displaying, using, or involving imagination or original ideas as well as routine skill or intellect, esp. in literature or art. c. Esp. of a financial or other strategy: ingenious, esp. in a misleading way. 2. Providing the cause or occasion of, productive of.

[1]For readability, some definitions are edited slightly to standardise formats and remove etymological/grammatical annotations.

create: 1.a. Said of the divine agent: To bring into being, cause to exist; esp. to produce where nothing was before, 'to form out of nothing'. b. with complemental extension. 2. To make, form, constitute, or bring into legal existence (an institution, condition, action, mental product, or form, not existing before). Sometimes of material works. 3. To constitute (a personage of rank or dignity); to invest with rank, title, etc. 4. To cause, occasion, produce, give rise to (a condition or set of circumstances).

The Penguin English Dictionary 2nd ed. (1969) p. 174:

creativity: creative power or faculty; ability to create

creative: having power to create; related to process of creation; constructive, original, producing an essentially new product; produced by original intellectual or artistic effort

create: make out of nothing, bestow existence on; cause, bring about; produce or make something new or original; confer new rank etc on; (theat) be the first to act (a certain part); make a fuss

Webster's 3rd New International Dictionary (1961) p. 532:

creativity: the quality of being creative; ability to create

creative: 1. having the power or quality of creating; given to creation 2: PRODUCTIVE - used with 3: having the quality of something created rather than imitated or assembled; expressive of the maker; IMAGINATIVE

create: 1: to bring into existence; make out of nothing and for the first time 2: to cause to be or to produce by fiat or by mental, moral, or legal action 3: to cause or occasion - used of natural or physical causes and esp. of social and evolutionary or emergent forces 4a: to produce (as a work of art or of dramatic interpretation) along new or unconventional lines) b: to design (as a costume or dress)

Given the problems inherent in dictionary definitions of creativity, it is not surprising that a number of creativity researchers have set out to provide their own definitions of the concept. Some examples are:

> 'creativity is that process which results in a novel work that is accepted as tenable or useful or satisfying by a group at some point in time' [18, p. 218]
>
> 'Creativity is the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)' [16, p. 3]
>
> 'Creativity is the ability to come up with ideas or artefacts that are *new*, *surprising and valuable*' [19, p. 1]
>
> 'Creativity is the interaction among *aptitude, process, and environment* by which an individual or group produces a *perceptible product* that is both *novel and useful* as defined within a *social context*' [9, p. 90]
>
> 'Creativity: the generation of products or ideas that are both novel and appropriate' [11, p. 570]
>
> 'The word creativity is a noun naming the phenomenon in which a person communicates a new concept (which is the product).

> Mental activity (or mental process) is implicit in this definition, and of course no one could conceive of a person living or operating in a vacuum, so the term *press* is also implicit' [7, p. 305]

These more research-oriented definitions avoid the problems of self-reference and circularity noted for the dictionary entries given previously. However, whilst the definitions may provide somewhat deeper insight into the nature of creativity, the brevity of the definitions means that they still only succeed in providing shallow, summary accounts of the concept.

A multitude of different perspectives The problem of identifying and quantifying creativity exists across many disciplines. How creative is this person? Does this person have the creative abilities to boost my business? Is this pupil's story creative? Is this computational system an example of computational creativity? As a consequence, when attempts are made to define creativity, it is often from the perspective of a particular domain or research discipline. For example, psychometric tests for creativity such as [20,21] focus on *problem solving* and *divergent thinking* as key attributes of a creative person. In contrast, computational creativity research [22–25, for example] has historially placed emphasis on the *novelty* and *value* of creative products. Whilst there is some consensus across academic fields, **for example novelty and value are typically recognised as necessary (but arguably not sufficient) components of creativity [26],** the differing emphases contribute to variations in the interpretation of creativity. These variations affect consistency across creativity research in different disciplines and potentially hinder interdisciplinary collaborations and cross-application of findings.

Several competing interpretations of creativity exist in the literature. Sometimes these differences of opinion do not need to be directly resolved but can be included alongside each other. Examples include whether creativity is centred around mental processes [19,27] or embodied and situated in an interactive environment [28,29]. Another example is whether creativity is domain-independent [30], or dependent on domain-specific context [31], or (as both Plucker and Baer have concluded) a combination of both [12,32].

Other conflicts arise where a previously narrow view of creativity has been widened in perspective. To resolve the conflict, an inclusive, all-encompassing view of creativity should adopt the wider perspective and incorporate the narrower perspective. For example rather than focussing narrowly on creative *genius*, through the study of people with exceptional creative achievements [33,34, for example] emphasis has shifted to encompass the broader study of *everyday* creativity, with genius as a special case: the notion that everyone can be creative to some degree [35,36].

Similarly, researchers distinguish between *little-c* and *Big-C* creativity, or *psychological/P-creativity* and *historical/H-creativity* [19], adjusting their focus accordingly to make their research more manageable. This is particularly the case in computational creativity, where endowing the computer with elements of general, human knowledge and experience is a major challenge. Little-c creative or p-creative work is perceived as creative by the creator personally but may replicate existing work (unknown to the creator) so is not necessarily creative in a wider social context. This encompasses the concept of Big-C creativity or h-creativity, where the work makes a creative contribution both to the creator and to society. To be Big-C creative/h-creative is to be little-c creative/p-creative in a way which has not been done before by anyone.

The preceeding discussion indicates that creativity is a complex, multi-faceted concept that requires a broad and inclusive treatment. The *Four Ps* framework [7,18,37–39] ensures we pay attention to four key aspects of creativity:

Person/Producer: The individual that is creative

Process: What the creative individual does to be creative

Product: What is produced as a result of the creative process

Press: The environment in which creative activity takes place

This framework presents creativity in a broader context, making our understanding of the concept more generally applicable and less specific to a domain or academic discipline. In contrast, models of the creative process [33,34,40], tests of people's creativity [21,41,42] or tests based on creative artefact generation [25,43] are useful only within a limited sphere. Jordanous [39] has contextualised the Four Ps in a computational context, referring to the creative Producer (person or computational agent) carrying out Processes within the environmental context of a Press, to create computational Products.

The challenges of modelling creativity

Creativity can be seen as an essentially contested concept [44]: it is subjective, abstract and can be interpreted in a variety of acceptable ways, such that a fixed 'proper general use' is elusive [44, p. 167]. Gallie defines an essentially contested concept through several features: being internally complex in nature, but amenable to being broken down into identifiable constituent elements of varying relative importance, and dependent on a number of factors such as context and individual preference. Although there may be consensus on the meaning of such concepts in very general terms, they may defy precise interpretation. There is not a single agreed instantiation, but instead many reasonable possibilities, influenced by changing circumstances and contexts. It is more productive to acknowledge that these different interpretations exist and refer to 'the respective contributions of its various parts or features' [44, p. 172], rather than to argue for a single interpretation. Thus, different types of creativity manifest themselves in different ways while sharing certain characteristics (not necessarily the same across all creative instances). This is what Wittgenstein refers to as 'family resemblances' [14]:

> [On discussing the example of what a 'game' is] 'we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail. ... I can think of no better expression to characterize these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. And I shall say: "games" form a family.' [14, Part 1, Paragraphs 66-67]

Similarly, with creativity, different manifestations of creativity are not all necessarily required to share the same common, core elements in order to be identified as part of the creativity 'family'. Rather, relationships between different manifestations reveal various shared characteristics that emerge in a similar way to Wittgenstein's 'family resemblances' in language. We need to identify what those family resemblances are in the case of creativity. To understand creativity, we can investigate what resemblances exist across different instantiations of the concept.

Wittgenstein [14] has argued that 'a clear view of the aim and functioning of the words' helps us 'dispers[e] the fog' that obscures a clear vision of the 'working of language' [14, Part 1, Paragraph 5]. To understand the use of a word, one must have background information and context. Wittgenstein gives the example of a chess piece,

which is introduced to someone as a 'king' [14, paragraph 31]: to understand this usage the person must already know the rules of chess, or must at least know what it means to have a piece in a game. To Wittgenstein, the semantics of words and statements are determined by how we use them, grounded in rules set by our habitual use of a word and our shared consensual practices, rather than being fixed by static, pre-assigned meanings.

Linguistics research advocates that the meaning of a word is dependent on the context it is used in [45]. In particular, Lakoff has argued that the study of language helps reveal how people think [13, 46]. Words used frequently in discussions of the nature of a concept provide the context for the commonly understood meaning of that concept, as has been shown in various corpus linguistics contributions [47–50].

The key principle emerging across these present discussions is that the meaning of words like creativity can be modelled by identifying different aspects that collectively contribute to the meaning of the concept of creativity.

The need for a clearer, multi-perspective understanding of creativity is evident, but remains to be addressed. There is a large quantity of material contributory to a satisfactory model of creativity and a number of key contributions have been discussed during this section. What must be done now is to marshal this assortment of material and to unify different perspectives where possible, in order to avoid the disciplinary 'blinkers' or compartmentalisation that is often seen in creativity research [11]. In approaching the semantic representation of subjective and multi-faceted concepts, some useful guidance is offered through philosophical reflections on the meaning of such concepts.

Methods

Our approach makes use of an empirical study and analysis of the language used to talk about creativity in order to gather and collate knowledge about the concept. In addition, following from the observations above, a *confluence approach* to creativity is adopted [16, 26, 51]. This works on the principle that creativity results from several components converging and goes on to examine what these components are. Taking this approach in conjunction with the application of tools from computational linguistics and statistical analysis allows a wider disciplinary spectrum of perspectives on creativity to be captured than has previously been attempted. This is achieved by breaking down the whole into smaller and more tractable constituent parts identified through a broad cross-disciplinary examination of creativity research.

Tools from natural language processing and statistical analysis are used to identify words that appear to be highly associated with dimensions of creativity, as represented in a sample of academic papers on the topic. A key innovation is the use of a statistical measure of lexical similarity, which allows the words to be clustered into coherent and semantically-related groups. Clustering reveals a number of common themes or factors of creativity, allowing the identification of a set of fourteen components that serve as building blocks for creativity.

Corpus data

A sample of academic papers discussing the nature of creativity was assembled as a *creativity corpus* in 2010. This creativity corpus consisted of 30 papers examining creativity from various academic stand-points ranging from psychological studies to computational models.

Creativity corpus:
a collection of thirty academic papers which explicitly discuss the nature of creativity.

The 30 papers selected for the creativity corpus are listed in Fig. 1a.

The search strategy for identifying papers for the corpus was based around a literature search for the term 'creativity' on the academic database *Scopus*. This literature search was supplemented with other influential papers which may have not appeared in a Scopus search; for example, a computer science conference paper on cognitive models of creativity.[2]. Paper selection for the creativity corpus was governed by inclusion criteria based on measuring the influence of a paper and coverage of a wide range of years and academic disciplines. The inclusion criteria are as follows, listed in descending order of precedence:

- **Papers must have, as their primary focus, discussion of the nature of creativity.**
- **Papers should be considered particularly influential. Influence was generally measured objectively, in terms of the number of times a paper had been cited by other academic authors. However, for papers published in recent years and which had consequently had little time to accrue citations, selection was based instead on a subjective judgement of influence grounded in a knowledge of the field.**
- **Papers selected should, as far as reasonable, represent a cross-section of years over the range 1950-2009. [The corpus was compiled in 2010.] 1950 was chosen as a starting point in recognition of the effect of J. P. Guilford's presidential address to the American Psychological Association [20], which examined contemporary creativity research (or more specifically, the lack of thereof). His talk was highly influential in encouraging more creativity research activity [10].**
- **Papers selected should, as far as reasonable, represent a cross-section of disciplines relevant to discussions of creativity. Fig. 1b. illustrates the disciplinary distribution of the corpus as it changes over the time period covered by the selected papers. This distribution is based on the *Scopus* database, which classifies journals under their main subject area(s) covered. We should acknowledge here though that while many disciplines include creative practice, often the focus is on application rather than in depth discussion of what creativity entails. Hence, while we sought to cover creativity from a broad range of perspectives, we also felt it was important not to compromise the focus of our corpus as a representation of key discussions about the nature of creativity.**

~~The papers were carefully chosen so as to cover a wide range of years (1950-2009) and academic disciplines. A paper was included if it was considered particularly influential, as measured by the number of times it had been cited by other academic authors. For papers published in recent years and which have consequently not yet accrued many citations, selection was based on subjective judgement of influence.~~

Exclusion criteria for this search were as follows:

[2]It should be noted that in Computer Science, a number of conferences carry as much or more publication weight as some journals in the field

- **Authors were only represented more than once in the corpus if the relevant papers were written from different perspectives. For example, Mark Runco's work is represented twice in the corpus, but covering two different topics relating to the nature of creativity (psychoeconomic approach to creativity; cognition and creativity). If the search process highlighted two or more papers with a shared author on the same or highly similar perspectives on creativity, then the more highly cited paper was chosen.**
- **Papers had to be written in English, as the language processing tools we were working with were for English language texts.**
- **Papers had to be available in a format that enabled us easily to extract plain text (this excluded books or book chapters).**

The creativity corpus is relatively small and necessarily selective in terms of the papers that are included. As such it constitutes just a small fraction of the many academic works on creativity that have been published in the last 60 or so years. Indeed, the 30 papers in the creativity corpus cannot be regarded as comprehensively representative of the wide range of academic positions on creativity that have been discussed in the literature over the decades. However, the goal of this work is not to present a fine-grained analysis of language use drawn from this complete literature, nor to provide a comprehensive lexicon or dictionary of creativity. Rather, the goal is to identify the broader ontological themes or factors that recur in our understanding of the concept of creativity. For this purpose, what is required is a sufficiently representative sample of the academic discourse on creativity. This sample can be used to identify the way in which word use reflects key themes or factors that persist across different perspectives.

Our objective is to identify what is distinctive in the language used to discuss creativity, in contrast to the language used to discuss other topics. As a basis for comparison, therefore, a further sample of 60 academic papers on topics unrelated to creativity – the *non-creativity corpus* – was assembled **alongside the creativity corpus, in 2010**.

Non-creativity corpus:
60 academic papers on topics unrelated to creativity, from the same range of academic disciplines and publication years as the creativity corpus papers.

The non-creativity corpus papers were selected by a literature search retrieving, for each paper in the creativity corpus, the two most-cited papers in the same academic discipline (as categorised by *Scopus*) and published in the same year, that did not contain any words with the prefix *creat* (i.e. *creativity*, *creative*, *creation*, and so on). **In other words, the criteria for inclusion in this second corpus were whether a paper was one of the two papers that was most highly cited at the time of the search (2010), in the same academic discipline, and published in the same year, as a paper in the creativity corpus, and that satisfied the exclusion criteria of not containing any words with the above mentioned prefixes.**

The non-creativity corpus is twice the size of the creativity corpus ($\approx$ 700,000 words and $\approx$ 300,000 words respectively), in acknowledgement of the fact that in general the set of academic papers on creativity is only a small subset of all academic papers. Both corpora are very small in comparison to corpora such as the British National Corpus, a

large (≈ 100M words) corpus of written and spoken English in general usage across a number of different contexts, and tiny in comparison to more recent web-derived text collections containing billions of words. There are, however, several benefits associated with using a corpus derived from specialist academic literature:

- Ease of locating relevant and appropriate papers: e.g. availability of tools to perform targeted literature searches, electronic publication of papers for download, tagging of paper content by keywords, citations in papers to other related papers.
- Ability to access timestamped textual materials over a range of decades.
- Publication of academic papers in an appropriate format for computational analysis: most papers that are available electronically are in formats such as PDF or HTML, which can be converted to text fairly easily.
- Availability of citation data as a measure of how influential a paper is on others: whilst not a perfect reflection of a paper's influence, citation data is often used for measuring the impact of a journal [52] or an individual researcher's output [53].
- Availability of provenance data, such as who wrote the paper and for what audience (from the disciplinary classification of the journal).

Some pre-processing was undertaken for each paper in both the creativity corpus and non-creativity corpus prior to analysis. A plain text file was generated for each paper, containing the full text of that paper. All journal headers and copyright notices were removed from each paper, as were the author names and affiliations, list of references and acknowledgements. All files were also checked for any non-ascii characters and anomalies that may have arisen during the creation of the text file.

Natural language processing

The corpus data was first pre-processed using the RASP natural language processing toolkit [54] in order to perform *lemmatisation* and *part-of-speech* tagging. Lemmatisation permits inflectional variants of a given word to be identified with a common 'dictonary headword' form or 'lemma'. For example, *performs*, *performed* and *performing* all occur in the creativity corpus as distinct morphological variants of the verb, *perform*. Intuitively, we would like to count each of these inflectional variants as an instance of the same word, rather than as separate and distinct lexical tokens. Lemmatisation software enables us to do this by mapping such variants to a cannonical lemma form. As a further refinement, each lemma was also mapped to lower case to ensure that capitalised word forms (e.g. *Novel*) were not counted separately from their non-capitalised forms (*novel*).[3]

Each word was assigned a part-of-speech tag identifying its grammatical category (i.e. whether the word was a noun, verb, preposition, etc.). Such tagging is useful because it allows us to distinguish between different grammatical uses of a common orthographic form. For example, the use of *novel* as a noun in *a good novel* can be properly differentiated from its use as an adjective in *a novel idea*. The data was further simplified and filtered so that only words of the four 'major' categories (i.e. noun, verb, adjective and adverb) were represented. Note that the major categories bear the semantic content of the papers making up the creativity corpus. They may be

[3]While this can result in confusion between proper names and common nouns (e.g. *Apple* v. *apple*), it is not considered that the resulting level of 'noise' in the data is likely to adversely affect the results of the analysis.

distinguished from minor categories or 'function words', such as pronouns (*something*, *itself*) prepositions (e.g. *upon*, *by*) conjunctions (*but*, *or*) and quantifiers (e.g. *many*, *more*). Because such words have little independent semantic content, they are of limited interest for the present study and may be removed from the data.

Following processing with RASP, a list of words found in the creativity corpus, together with their frequency counts was generated. The non-creativity corpus was pre-processed in the same way and a corresponding list of words and frequencies also generated.

Identifying words associated with creativity

The word frequency data derived from the two corpora was used to establish which words occur significantly more often in the creativity corpus than in the non-creativity corpus. This in turn can be regarded as providing evidence of which words are salient to the definition of creativity. Salient words were identified using the log-likelihood ratio (also referred to as the G^2 or G-squared statistic), which is a measure of how well observed data fit a model or expected distribution [47–49, 55]. It provides an alternative to Pearson's chi-squared (χ^2) test and has been advocated as the more appropriate measure of the two for corpus analysis as it does not rely on the (unjustifiable) assumption of normality in word distribution [47, 49, 55]. This is a particular issue when analysing smaller corpora, such as those used in the present work (see Fig. 1). The log likelihood ratio statistic is more accurate in its treatment of infrequent words in the data, which often hold useful information. By contrast, the χ^2 statistic tends to under-emphasise such outliers at the expense of very frequently occurring data points.

Our use of the log-likelihood ratio follows that of Rayson and Garside [48]. Given two corpora (in our case, the creativity corpus *cc* and the non-creativity corpus *nc*) the log-likelihood score for a given word is calculated as shown in equation (1) below:

$$LL = O_{cc} \ln(\frac{O_{cc}}{E_{cc}}) + O_{nc} \ln(\frac{O_{nc}}{E_{nc}}) \quad (1)$$

where O_{cc} (O_{nc}) is the observed frequence of the word in *cc* (*nc*) and similarly E_{cc} (E_{nc}) is its expected frequency. The expected frequency E_{cc} is given by:

$$E_{cc} = \frac{N_{cc} \times (O_{cc} + O_{nc})}{N_{cc} + N_{nc}} \quad (2)$$

where N_{cc} denotes the total number of words in corpus *cc* (i.e. the sum of the frequencies of all words drawn from corpus *cc*). The expected frequency E_{nc} is defined in a way analogous to Equation 2.

As computed above, the log-likelihood ratio measures the extent to which the distribution of a given word deviates from what might be expected if its distribution is not corpus dependent. The higher the log likelihood ratio score for a given word, the greater the deviation from what is expected. It should be noted however, that the statistic tells us only that the observed distribution of a word in the two corpora is unexpected (and to what extent). It does not tell us whether the word is more or less frequent than expected in the creativity corpus. To identify words significantly associated with creativity therefore, it was necessary to select just those words with observed counts higher than that expected in the creativity corpus. It should perhaps be further noted that the resulting words may be either positively or negatively connoted with respect to creativity. In practice this is not a problem, as the significance of a given word lies in its semantic connection to creativity, not in its sentiment or affect. Affect is taken into account as part of the later manual examination of the data used to identify components of creativity.

The results of the calculations were filtered to remove any words with a log-likelihood score less than 10.83, representing a chi-squared significance value for p=0.001 (one degree of freedom). In this way, the filtering process reduced the set of candidate words to just those that appear to occur significantly more often than expected in the creativity corpus. To avoid extremely infrequent words disproportionately affecting the data, any word occurring fewer than five times was also removed from the data. Finally, the words were inspected to remove any 'spurious' items such as proper nouns or misclassified or odd character sequences. This resulted in a total of 694 *creativity words*: a collection of 389 nouns, 205 adjectives, 72 verbs and 28 adverbs that occurred significantly more often than expected in the creativity corpus. Table 1 gives the top 20 results of these calculations.

Identifying components of creativity

It is important to note that our objective is to identify key themes in the lexical data, not to induce a comprehensive terminology of creativity. Despite the relatively small size of the corpora used, the resulting set of 694 creativity words is sufficiently rich for this purpose, but is still somewhat large to work with in its raw form. In previous, related work [56] an attempt was made to identify key components by manually clustering creativity words by inspection of the raw data. In practice, this proved laborious and made it impossible systematically to consider all of the identified words. It also raised issues of subjectivity and experimenter bias. These problems are addressed here, at least in part, by automatically clustering the words according to a statistical measure of *distributional similarity* [57]. The more manageable collection of clusters may then be examined to identify key components or dimensions of creativity.

~~The creativity words were clustered according a statistical measure of *distributional similarity*.~~ The intuition underlying distributional measures of similarity derives from the *distributional hypothesis* due to Harris [58]. This hypothesis states that similarity of distribution correlates with similarity of meaning: two words that tend to appear in similar linguistic contexts will tend to have similar meanings. The notion of linguistic context here is not fixed and might plausibly be modelled in a variety of different ways. For example, two words might be considered to inhabit the same context if they appear in the same document or the same sentence or if they stand in the same grammatical relationship to some other word (e.g. both occur as *subject* of a particular verb or *modifier* of a given noun). In practice it has been shown that modelling distribution in terms of grammatical relations leads to a tighter correlation between distributional similarity and closeness of meaning [59].

In the present work, grammatical relations are used to represent linguistic context and distributional similarity is measured as a function of the number of relations that two words share. To illustrate, evidence that the words *concept* and *idea* are similar in meaning might be provided by occurrences such as the following:

(1) the *concept/idea* involves	(subject of verb 'involve')
(2) applied the *concept/idea*	(object of verb 'apply')
(3) the basic *concept/idea*	(modified by adjective 'basic')

Grammatical relations were obtained from an analysis of the written portion of the British National Corpus [60], which had previously been processed using the RASP toolkit [54] in order to extract them. Using this data, each word in the creativity corpus was associated with a list of all of the grammatical relations with which it occurred, together with their corresponding counts of occurrence.[4] A potential difficulty with

[4]Not all of the grammatical relation information output by RASP was used to calculate distributional similarity. In practice, just the subject, object and modifier relation types are used as these tend to give the best results [61].

obtaining word similarity data based on the BNC (i.e. using data from sources of everyday usage of English, rather than from more specialist sources) would arise if the majority of the creativity words were used with distinctive or technical senses within the creativity corpus. From inspection and from knowledge of creativity literature, however, this situation was found to be unlikely. While some narrowly specialised usage may be present to some small degree in the set of creativity words, most words retain general senses as reflected in the wider BNC data set. An advantage of using the BNC is that its size increases the chances of a comprehensive coverage of the general senses of each word of interest.

Distributional similarity of two words is measured in terms of the similarity of their associated lists of grammatical relations. A variety of different methods for calculating distributional similarity have been investigated in the literature, including standard techniques such as the cosine measure [62, for example]. The present work adopts an information-theoretic measure due to Lin [57], which has been widely used in language processing applications as a means of automatically discovering semantic relationships between words. In comparison to other similarity measures it has been shown to perform particularly well as a means of identifying near-synonyms [63, 64].

Similarity scores were calculated between all pairs of creativity words of the same grammatical category. That is, scores were obtained separately for pairs of nouns, verbs, adjectives and adverbs. For a given set of words, word pair similarity data calculated in this way can be conveniently visualised as an edge-weighted graph, where nodes correspond to words and edges are weighted by similarity scores (for any score > 0), as in Fig. 2.

Graphical representations of the similarity data like that shown in Fig. 2 provide a useful basis for analysing the creativity words and identifying recurring themes or components of creativity. Two complementary methods for identifying key components of the data were adopted:

Clustering: The graph clustering software *Chinese Whispers* [65] was used to automatically identify word clusters (groups of closely interconnected words) in the dataset. This algorithm uses an iterative process to group together graph nodes that are located close to each other. By grouping words with similar meanings, the number of data items was effectively reduced and themes in the data could be recognised more readily from each distinct cluster. A sample of some of the resulting clusters can be seen in Fig. 3.

Inspection: To focus on the words most closely related to creativity, the top twenty creativity words (i.e. the twenty words with the highest log likelihood scores) were selected. Each word was then visualised as the root node of its own individual subgraph using the graph drawing software *GraphViz* [66]. In order to reduce the amount of data to be examined, similarity scores were discarded if they fell below a threshold value (adjusted manually for each graph to highlight the most strongly connected words). This made the size and complexity of the graphs smaller and therefore easier to inspect and analyse visually. Fig. 4 illustrates, in diagram form, the process of using manual inspection to identify components.

As part of the manual inspection process, candidate components were further considered in terms of the *Four Ps* of creativity [7, 10, 37, 67] described earlier in this paper. This additional analysis provided a means of identifying alternative perspectives and revealing subtle (but still important) aspects of creativity. For example, *novelty* is commonly associated with the results of creative behaviour (*product*): how novel is the artefact or idea that has been produced? However, we could similarly recognise as creative an approach to a task (*process*) that does things in a novel and different way.

Also, if a product is new in a particular environment (*press*), then it may well be regarded as creative to those in that environment. Viewing *novelty* from the perspectives of *product*, *process* and *press* uncovers these subtle and interlinked distinctions.

Results and Discussion

Components of creativity

From the analysis steps described in the previous section it was possible to extract a set of fourteen key components of creativity. These components are summarised in Fig. 5 and are presented in more detail below. The components contribute collectively to the overall concept and may be regarded as providing an *ontology of creativity*. It is important to note, however that the fourteen components do not constitute a set of necessary and sufficient conditions for creativity, in all its possible manifestations. There are two reasons for this. Firstly, some of the components we have identified appear to be logically inconsistent with others in the set. Consider for example the apparent need for autonomous, independent behaviour identified in *Independence and Freedom* and contrast this with the requirement for social interaction implied by *Social Interaction and Communication*. Secondly, of course, creativity also manifests itself in rather different ways across different domains [12] and components will vary in importance, according to the requirements of a particular domain. As an illustration of this second point, creative behaviour in mathematical reasoning has more focus on finding a correct solution to a problem than is the case for creative behaviour in, say, musical improvisation [2, 68].

The following set of fourteen components is therefore presented as a collection of dimensions – attributes, abilities and behaviours, etc. – which contribute to our understanding of creativity. The components should be treated as *building blocks* for creativity that may be arranged in different ways and with different emphases to suit different modelling purposes. The analysis of creativity in terms of the dimensions should be informative for a human audience and provide a basis for machine-understanding of the concept. Each component is presented here with a brief explanation or gloss. These explanations will later be used for part of the semantic content in the creativity ontology.

Active Involvement and Persistence:

Being actively involved; reacting to and having a deliberate effect on the creative process.

The tenacity to persist with the creative process throughout, even during problematic points.

Dealing with Uncertainty:

Coping with incomplete, missing, inconsistent, contradictory, ambiguous and/or uncertain information. Element of risk and chance - no guarantee that information problems will be resolved.

Not relying on every step of the process to be specified in detail; perhaps even avoiding routine or pre-existing methods and solutions.

Domain Competence:

Domain-specific intelligence, knowledge, talent, skills, experience and expertise. Knowing a domain well enough to be equipped to recognise gaps, needs or problems that need solving and to generate, validate, develop and promote new ideas in that domain.

General Intellectual Ability:
General intelligence and IQ.
Good mental capacity.

Generation of Results:
Working towards some end target, goal, or result.
Producing something (tangible or intangible) that previously did not exist.

Independence and Freedom:
Working independently with autonomy over actions and decisions.
Freedom to work without being bound to pre-existing solutions, processes or biases; perhaps challenging cultural or domain norms.

Intention and Emotional Involvement:
Personal and emotional investment, immersion, self-expression and involvement in the creative process.
The intention and desire to be creative: creativity is its own reward, a positive process giving fulfilment and enjoyment.

Originality:
Novelty and originality; a new product, or doing something in a new way; seeing new links and relations between previously unassociated concepts.
Results that are unpredictable, unexpected, surprising, unusual, out of the ordinary.

Progression and Development:
Movement, advancement, evolution and development during a process.
Whilst progress may or may not be linear, and an actual end goal may be only loosely specified (if at all), the entire process should represent some progress in a particular domain or task.

Social Interaction and Communication:
Communicating and promoting work to others in a persuasive and positive manner.
Mutual influence, feedback, sharing and collaboration between society and individual.

Spontaneity/Subconscious Processing:
No need to be in control of the whole process; thoughts and activities may inform the process subconsciously without being inaccessible for conscious analysis, or may receive less attention than others.
Being able to react quickly and spontaneously when appropriate, without needing to spend too much time thinking about the options.

Thinking and Evaluation:
Consciously evaluating several options to recognise potential value in each and identify the best option, using reasoning and good judgement.
Proactively selecting a decided choice from possible options, without allowing the process to stagnate under indecision.

Value:
Making a useful contribution that is valued by others and recognised as an achievement and influential advancement; perceived as special, 'not just something anybody would have done'.
The end product is relevant and appropriate to the domain being worked in.

Variety, Divergence and Experimentation:

Generating a variety of different ideas to compare and choose from, with the flexibility to be open to several perspectives and to experiment and try different options out without bias.

Multi-tasking during the creative process.

Implementing a machine-readable ontology of creativity

The fourteen components provide a fuller and clearer account of the constituent parts of the concept of creativity. An important aim of the current work is to make the components available as a resource for other researchers in computational creativity and to provide a basis for the automated evaluation of creative systems. As a step in this direction, the components have been expressed in an open, machine-readable form within the Semantic Web. In this way, the characterization of the components benefits from and is enriched by concepts that are already represented within the Semantic Web.

In particular, the components are linked to the data in WordNet [69], a large lexical database of English that has recently been made available as a Semantic Web ontology [70]. In WordNet, words are grouped by sense and interlinked by lexical and conceptual relations. Note that, although the WordNet definition of the word such as 'creativity' is brief ('the ability to create'), its utility lies in how it is linked to various concepts, such as its sense, hyponyms, type, 'gloss' (brief definition) and other related concepts. Each creativity component relates to a cluster of keywords from the original set of 694 creativity words. Following Linked Data principles, each can therefore be linked across the Semantic Web to an appropriate set of concepts from WordNet. In this way, associated semantic information is provided for each component.

The resulting encoding can be visualised as a graph, as shown in Fig. 6. The data has also been published under an Open Data Commons Public Domain Dedication and Licence (PDDL) [71] at:

`http://purl.org/creativity/ontology`

The concept labelled *Creativity* has the unique URI:

`http://purl.org/creativity/ontology#Creativity`

Any Linked Data that needs to refer to the concept can use this identifier.

Evaluation

From a practical stand-point, the current work is part of an overarching project engaged with the question of the evaluation of creativity, particularly computational creativity [72]. It is clear that a rigorous and comparative evaluation process needs clear standards to use as guidelines or benchmarks [10, 15].

The components of creativity in this paper have been employed in two case studies to evaluate the creativity of computational systems [68, 73, 74]. In these case studies, evaluation was carried out using the three step approach of the Standardised Procedure for Evaluating Creative Systems (SPECS) [72]:

1. Identify a definition of creativity that your system should satisfy to be considered creative:
   (a) What does it mean to be creative in a general context, independent of any domain specifics?

(b) What aspects of creativity are particularly important in the domain your system works in (and what aspects of creativity are less important in that domain)?

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.

In both case studies, the components of creativity were chosen as the way of characterising creativity for step 1a of SPECS, and were weighted according to their importance and relevance for creativity in the creative domains under study for each case study (step 1b of SPECS). Each component was treated as one standard to be used to evaluate the creativity of the creative systems in the case studies (step 2 of SPECS). Each case study system was then tested against each component using feedback provided by judges (step 3 of SPECS), resulting in a detailed set of evaluative feedback on the creativity of each system in the case studies.

Case Study 1 [72, 73] evaluated the creativity of three different computational musical improvisation systems [72]. Case Study 2 used the components of creativity in an evaluation scenario where information and time was limited for evaluation, to simulate the forming of first impressions and snapshot judgements of the creativeness of a given computational creativity system [73, 74].

The resulting component-based evaluation yielded detailed information about creative strengths and weaknesses of the systems under investigation, highlighting those components where a system performs strongly. Crucially, the evaluation feedback also highlighted areas where a given system performed poorly. For example, in the musical improvisation study, Case Study 1, we found that, in general, creativity could be improved most by improving performance in *Social Interaction and Communication*, *Intention and Emotional Involvement* and *Domain Competence* (the three components found to be most important for creativity in musical improvisation). Similarly, it is useful to be able to quickly obtain formative feedback on strengths and weaknesses in time-limited scenarios such as that replicated in Case Study 2 during the development of creative systems (when ongoing evaluation of progress ideally needs to be both timely and time-efficient). Insight can then be obtained on where future development effort is best spent.

The results obtained above were compared with those from applying other evaluation models and with surveys of people's opinions, where people were asked how creative they thought each system was. There was general agreement between evaluation approaches on the most and least creative systems. The approaches differed in the formative feedback they provided, particularly for identifying strengths of the system at being creative, and weaknesses of the system to be improved. The model of creativity offered in this paper gave the most detailed feedback, but required most information to be collected.

To support the usefulness of having the components as a tangible characterisation of creativity, an interesting finding was made as part of the first case study, in a separate evaluation carried out: asking for people's opinions on how creative the musical improvisation systems were. A striking observation was that a number of participants called for the word "creativity" to be defined before they felt comfortable with the task and confident in evaluating creativity in this setting, even though participants reported feeling generally positive or at least neutral towards the concept of computational creativity. This challenges the generally held view that people have a common-sense working definition of creativity, at least in the context of judgement and evaluation. A representation of creativity is useful to:

1. establish what it means for something to be deemed creative; and

2. identify appropriate evaluation standards that replicate typical human opinion on how creative something is or in comparing two or more creative systems.

Conclusions and directions for future work

This paper has described the methods used to identify a set of *components of creativity* using corpus-based, statistical language processing techniques. The motivation for the work is the need for a shared, comprehensive and multi-perspective model of creativity. Such a model should be of great value to researchers investigating the nature of creativity and in particular those concerned with the evaluation of creative practice. More broadly, the inter-disciplinary approach described here exemplifies a general approach to the investigation and representation of semantically fuzzy and essentially-contested concepts. For this reason, we expect that it will interest researchers investigating computational methods for analysing and representing other such concepts.

Rather than attempting to provide a unitary definition of creativity, our approach extracts common, underlying themes that transcend discipline or domain bias. Our point of departure is the observation that the vocabulary used in discussions of the nature of creativity may be analysed in order to throw light on our understanding of the concept and its key attributes. Using techniques from corpus linguistics and natural language processing (as described in the Methods section), key components of creativity have been identified. The results of this novel, empirical analysis (presented in the Results section) inform the development of an *ontology of creativity* comprising a set of fourteen distinct components (see Fig. 5). It is noted that each component makes a separate contribution to the overall meaning of the concept. At the same time, because creativity manifests itself in different ways across different domains [12], the individual components vary in importance and influence according to the requirements of a given domain. The components can be therefore be usefully thought of as 'building blocks' for the concept in its different manifestations. Taken together, the components make creativity more tractable to study and to evaluate.

The fourteen components provide a multi-perspective model of creativity that has been successfully applied in a comparative analysis and evaluation of computational creativity systems [72–74] (see the Discussions section). The outcome of the evaluation process provides relatively fine-grained information about the creative strengths of a given system. This information in turn evidences ways in which a system could be considered creative. In addition, evaluation based on the components is able to highlight areas of weakness. These can be used to inform future work aimed at further developing a system's creative potential.

The components have been published in an open, machine-readable format, making them freely available to the research community. This has a number of implications. First, the set of components may be readily elaborated, extended or amended by other researchers investigating the concept of creativity. Second, the machine-readable format facilitates the development of creativity-aware applications, based on the components. Such applications might be developed to support manual evaluation of creative practice or as a significant step towards the development of methods for automated evaluation.

The problem of developing automated evaluation has elsewhere been described as 'the Achilles' heel of AI research on creativity' [75]. An intriguing possibility that we are currently exploring is to further exploit language processing techniques to perform evaluation based on textual reviews, descriptions of system performance, or social media interactions [76]. Such an approach would be analogous to the way sentiment analysis techniques are now in common use to evaluate attitude and opinion based on reviews of

products or services [77]. This is a fascinating direction for future work, with great potential for real progress towards tackling computational creativity's 'Achilles' heel'.

Supporting Information

S1a Fig

The 30 papers that make up the Creativity Corpus

S1 Fig

The 30 papers that make up the Creativity Corpus

S1b Fig

Representation of the disciplinary breakdown of the Creativity Corpus by time period. Disciplines are as specified for the paper's journal, by the academic database *Scopus*. Note that Scopus may classify a journal under more than one discipline.

S3 Fig

Sample of clusters produced by the Chinese Whispers clustering step.

S4 Fig

Illustration of the process of using manual inspection for further clustering.

S5 Fig

The fourteen key components of creativity identified through an analysis of the word clusters

S6 Fig

The ontology of Creativity, in graph form.

S1 Table

The top 20 results of the log-likelihood ratio (LLR) calculations. A significant LLR score at p=0.001 is 10.83. N.B. POS=Part Of Speech: N=noun, J=adjective, V=verb, R=adverb.

Acknowledgments

We would like to acknowledge Nick Collins and Chris Thornton for their helpful comments during this work.

References

1. Colton S, Wiggins GA. Computational Creativity: The Final Frontier? In: Proceedings of 20th European Conference on Artificial Intelligence (ECAI). Montpellier, France; 2012. p. 21–26.
2. Colton S. Creativity versus the Perception of Creativity in Computational Systems. In: Proceedings of AAAI Symposium on Creative Systems; 2008. p. 14–20.
3. Widmer G, Flossmann S, Grachten M. YQX Plays Chopin. AI Magazine. 2009;30(3):35–48.
4. León C, Gervás P. The Role of Evaluation-Driven Rejection in the Successful Exploration of a Conceptual Space of Stories. Minds and Machines. 2010;20(4):615–634.
5. Pérez y Pérez R. MEXICA: A Computer Model of Creativity in Writing. University of Sussex. Brighton, UK; 1999.
6. Cardoso A, Veale T, Wiggins GA. Converging on the Divergent: The History (and Future) of the International Joint Workshops in Computational Creativity. AI Magazine. 2009;30(3):15–22.
7. Rhodes M. An analysis of creativity. Phi Delta Kappan. 1961;42(7):305–310.
8. Torrance EP. Scientific Views of Creativity and Factors Affecting its Growth. In: Kagan J, editor. Creativity and Learning. Boston: Beacon Press; 1967. p. 73–91.
9. Plucker JA, Beghetto RA, Dow GT. Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research. Educational Psychologist. 2004;39(2):83–96.
10. Kaufman JC. Creativity 101. The Psych 101 series. New York: Springer; 2009.
11. Hennessey BA, Amabile TM. Creativity. Annual Review of Psychology. 2010;61:569–598.
12. Plucker JA, Beghetto RA. Why Creativity is Domain General, Why it Looks Domain Specific, and why the Distinction Doesn't Matter. In: Sternberg RJ, Grigorenko EL, Singer JL, editors. Creativity: From Potential to Realization. Washington, DC: American Psychological Association; 2004. p. 153–167.
13. Lakoff G. Women, Fire and Dangerous things: What Categories reveal about the mind. Chicago, IL: University of Chicago Press; 1987.
14. Wittgenstein L. Philosophical Investigations, eds. Anscombe, G. E. M. and Rhees, R. and Von Wright, G. H. 2nd ed. Oxford, UK: Basil Blackwell; 1958.
15. Torrance EP. The Nature of Creativity as Manifest in its testing. In: Sternberg RJ, editor. The Nature of Creativity. Cambridge, UK: Cambridge University Press; 1988. p. 43–75.
16. Sternberg RJ, Lubart TI. The Concept of Creativity: Prospects and Paradigms. In: Sternberg RJ, editor. Handbook of Creativity. Cambridge, UK: Cambridge University Press; 1999. p. 3–15.

17. Vartanian O. Toward a Cumulative Psychological Science of Aesthetics, Creativity, and the Arts. Psychology of Aesthetics, Creativity, and the Arts. 2014;8(1):15–17.

18. Stein MI. A Transactional Approach to Creativity. In: Taylor CW, Barron F, editors. Scientific Creativity: Its Recognition and Development. New York: John Wiley & Sons; 1963. p. 217–227.

19. Boden MA. The creative mind: Myths and mechanisms. 2nd ed. London, UK: Routledge; 2004.

20. Guilford JP. Creativity. American Psychologist. 1950;5:444–454.

21. Torrance EP. Torrance Tests of Creative Thinking. Bensenville, IL: Scholastic Testing Service; 1974.

22. Pease A, Winterstein D, Colton S. Evaluating Machine Creativity. In: Proceedings of Workshop Program of ICCBR-Creative Systems: Approaches to Creativity in AI and Cognitive Science; 2001. p. 129–137.

23. Wiggins GA. Searching for computational creativity. New Generation Computing. 2006;24(3):209–222.

24. Peinado F, Gervas P. Evaluation of automatic generation of basic stories. New Generation Computing. 2006;24(3):289–302.

25. Ritchie G. Some Empirical Criteria for Attributing Creativity to a Computer Program. Minds and Machines. 2007;17:67–99.

26. Mayer RE. Fifty Years of Creativity Research. In: Sternberg RJ, editor. Handbook of Creativity. Cambridge, UK: Cambridge University Press; 1999. p. 449–460.

27. Dietrich A, Kanso R. A review of EEG, ERP, and Neuroimaging Studies of Creativity and Insight. Psychological Bulletin. 2010;136(5):822–848.

28. McCormack J. Creative Ecosystems. In: Proceedings of the 4th International Joint Workshop on Computational Creativity. London, UK; 2007. p. 129–136.

29. Sosa R, Gero J, Jennings K. Growing and Destroying the Worth of Ideas. In: Proceedings of the 7th ACM Creativity and Cognition conference. Berkeley, California; 2009. p. 295–304.

30. Plucker JA. Beware of Simple Conclusions: The Case for Content Generality of Creativity. Creativity Research Journal. 1998;11(2):179–182.

31. Baer J. The Case for Domain Specificity of Creativity. Creativity Research Journal. 1998;11(2):173–177.

32. Baer J. Is Creativity Domain-Specific? In: Kaufman JC, Sternberg RJ, editors. The Cambridge Handbook of Creativity. New York, NY: Cambridge University Press; 2010. p. 321–341.

33. Poincaré H. Mathematical Creation. In: The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method.. vol. Science and Method [Original French version published 1908, Authorized translation by George Bruce Halsted]. New York: The Science Press; 1929. p. 383–394.

34. Hadamard J. An Essay on the Psychology of Invention in the Mathematical Field. Princeton, NJ: Princeton University Press; 1945.

35. Weisberg RW. Problem Solving and Creativity. In: Sternberg RJ, editor. The Nature of Creativity. Cambridge, UK: Cambridge University Press; 1988. .

36. Bryan-Kinns N. Everyday Creativity. In: Proceedings of the 7th ACM conference on Creativity and Cognition. Berkeley, California; 2009. .

37. Mooney RL. A Conceptual Model for Integrating Four Approaches to the Identification of Creative Talent. In: Taylor CW, Barron F, editors. Scientific Creativity: Its Recognition and Development. New York: John Wiley & Sons; 1963. p. 331–340.

38. Odena O, Welch G. A Generative Model of Teachers' Thinking on Musical Creativity. Psychology of Music. 2009;37(4):416–442.

39. Jordanous A. Four PPPPerspectives on Computational Creativity in theory and in practice. Connection Science. 2016;tbc(tbc).

40. Wallas G. The Art of Thought. 1st ed. London, UK: Jonathan Cape; 1926.

41. Goldman RJ. The Minnesota Tests of Creative Thinking. Educational Research. 1964;7(1):3–14.

42. Guilford JP. The nature of human intelligence. New York, NY: McGraw-Hill; 1967.

43. Amabile TM. Creativity in context. Boulder, Colorado: Westview Press; 1996.

44. Gallie WB. Essentially Contested Concepts. Proceedings of the Aristotelian Society. 1956;56:167–198. Available from: http://www.jstor.org/stable/4544562.

45. Firth JR. A synopsis of linguistic theory, 1930-1955. Studies in Linguistic Analysis. 1957;p. 1–32.

46. Lakoff G, Johnson M. Metaphors we live by. Chicago, IL: University of Chicago Press; 1980.

47. Oakes MP. Statistics for Corpus Linguistics. Edinburgh, UK: Edinburgh University Press; 1998.

48. Rayson P, Garside R. Comparing Corpora using Frequency Profiling. In: Proceedings of ACL Workshop on Comparing Corpora. Hong Kong; 2000. .

49. Kilgarriff A. Comparing Corpora. International Journal of Corpus Linguistics. 2001;6(1):97–133.

50. Kilgarriff A. Where to go if you would like to find out more about a word than the dictionary tells you. Macmillan English Dictionary Magazine. 2006;Issue 35 (Jan-Feb).

51. Ivcevic Z. Creativity Map: Toward the Next Generation of Theories of Creativity. Psychology of Aesthetics, Creativity, and the Arts. 2009;3(1):17–21.

52. Garfield E. Citation analysis as a tool in journal evaluation. Science. 1972;178(60):471–479.

53. Hirsch JE. An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(46):16569–16572.

54. Briscoe E, Carroll J, Watson R. The Second Release of the RASP System. In: Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. Sydney, Australia; 2006. .

55. Dunning T. Accurate Methods for the Statistics of Surprise and Coincidence. Computational Linguistics. 1993;19(1):61–74.

56. Jordanous A. Defining Creativity: Finding Keywords for Creativity Using Corpus Linguistics Techniques. In: Proceedings of the International Conference on Computational Creativity. Lisbon, Portugal; 2010. p. 278–287.

57. Lin D. An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning. Madison, WI; 1998. p. 296–304.

58. Harris Z. Mathematical Structures of Language. New York: Wiley; 1968.

59. Kilgarriff A, Yallop C. What's in a thesaurus. In: Proceedings of the Second Conference on Language Resources and Evaluation (LREC-00). Athens; 2000. p. 1371–1379.

60. Leech G. 100 million words of English: the British National Corpus (BNC). Language Research. 1992;28(1):1–13.

61. Weeds JE. Measures and Applications of Lexical Distributional Similarity. Informatics, University of Sussex. Brighton, UK; 2003.

62. Manning C, Schütze H. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press; 1999.

63. Weeds J, Weir D. Finding and evaluating nearest neighbours. In: Proceedings of the 2nd Conference of Corpus Linguistics. Lancaster, UK; 2003. .

64. McCarthy D, Navigli R. The English lexical substitution task. Language Resources and Evaluation: Special Issue on Computational Semantic Analysis of Language. 2009;43(3):139–159.

65. Biemann C. Chinese Whispers: an efficient graph clustering algorithm and its application to natural language processing problems. In: Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing. Morristown, NJ: Association for Computational Linguistics; 2006. p. 73–80.

66. GraphViz. GraphViz - Graph Visualization Software; 1988. *http* : *//www.graphviz.org/*, last accessed January 2013.

67. MacKinnon DW. Creativity: a Multi-Faceted Phenomenon. In: Roslansky JD, editor. Creativity: A Discussion at the Nobel Conference. Amsterdam, The Netherlands: North-Holland Publishing Company; 1970. p. 17–32.

68. Jordanous A, Keller B. What makes musical improvisation creative? Journal of Interdisciplinary Music Studies. 2012;6:151–175.

69. Fellbaum C, editor. WordNet: An electronic lexical database. Cambridge, MA: The MIT press; 1998.

70. RKBExplorer. wordnet.rkbexplorer.com; 2012. *http* : *//wordnet.rkbexplorer.com/*, last accessed January 2013.

71. Miller P, Styles R, Heath T. Open data commons, a license for open data. In: Proceedings of the WWW2008 Workshop on Linked Data on the Web. Beijing, China; 2008. .

72. Jordanous A. A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative. Cognitive Computation. 2012;4(3):246–279.

73. Jordanous A. Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application. University of Sussex. Brighton, UK; 2012.

74. Jordanous A. The longer term value of creativity judgements in computational creativity. In: Submitted to AISB Symposium on Computational Creativity (CC2016); 2016. .

75. Boden MA. Introduction [summary of Boden's keynote address to AISB'99]. In: AISB Quarterly - Special issue on AISB99: Creativity in the arts and sciences. vol. 102; 1999. p. 11.

76. Jordanous A, Allington D, Dueck B. Measuring cultural value using social network analysis: a case study on valuing electronic musicians. In: Proceedings of the Sixth International Conference on Computational Creativity June; 2015. p. 110.

77. Pang B, Lee L. Opinion mining and sentiment analysis. Foundation and Trends in Information Retrieval. 2008;2(1-2):1–135.