Creativity is what we say it is: constructing an ontology of creativity

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Anna Jordanous [a,*] and Bill Keller [b]
[a] *Centre for e-Research, Department of Digital Humanities, King's College London, 26-29 Drury Lane, London WC2B 5RL, UK*
E-mail: anna.jordanous@kcl.ac.uk
[b] *Department of Informatics, University of Sussex, Falmer, Brighton BN1 9RH, UK*
E-mail: billk@sussex.ac.uk

Abstract. This paper describes a novel application of cognitive science methods and approaches to the analysis of creativity and to the development of a machine-readable ontology of creativity. As a complex, ambiguous and multi-faceted concept, creativity has proved resistant to satisfactory and comprehensive definition, despite numerous attempts to provide one. Using techniques from the field of statistical natural language processing and cognitive modelling techniques, this paper describes the construction of a representation of creativity on the Semantic Web, built through an analysis of what is considered important when talking about creativity. Words were identified which appeared significantly often in connection with discussions of the concept. Using a measure of lexical similarity to cluster these words, a number of distinct themes emerged, which collectively contribute to a comprehensive and multi-perspective representation of creativity. The components provide an *ontology of creativity* in the form of a set of *building blocks* that collectively make this subjective concept more tractable, increasing semantic clarity and depth of information available on the concept while accommodating different instantiations of creative activity. The ontology has application to research into the nature of creativity in general and to the evaluation of creative practice, in particular. The provision of a machine readable conceptualisation of creativity also provides a small but significant step towards addressing the problem of automated evaluation, 'the Achilles' heel of AI research on creativity' [7]. From a broader perspective, building the creativity ontology illustrates how adoption of cognitive science methods enables the modelling and representation of highly subjective, semantically contestable concepts on the Semantic Web: a significant step forward for Semantic Web research.

Keywords: creativity, ontology of creativity, natural language processing, log likelihood ratio, lexical similarity

1. Introduction

Currently, most content on the Semantic Web is in the form of ontologies of 'things': semantically structured collections of factual or objective data on topics as diverse as people [21], places [25], music [58] or manuscripts [75]. The focus in Semantic Web research has generally been on reducing ambiguity in representations of these more structured domains and concepts, rather than directly tackling the representation and machine-readable representation of subjective, unstructured concepts in an ontology. However, cognitive science research and work on lexical resources such as WordNet [19,74] have laid foundations for more definitionally troublesome concepts to be considered in detail. The time is ripe for developing ontologies of subjective concepts such as creativity.

*Corresponding author. E-mail: anna.jordanous@kcl.ac.uk, Tel: +44 (0)20 7848 1988.

Creativity is a complex, multi-faceted concept encompassing many related aspects, abilities, properties and behaviours, as discussed in section 1.1. There have been many attempts to capture the nature of creativity in words; indeed the work described in this paper is based on thirty such examples (see section 2.1 and Figure 2.1). Yet despite many attempts to provide a workable and comprehensive definition of the concept, no real consensus has yet been reached on what exactly constitutes creativity. In the academic literature on creativity, many repeated 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 comprehensive, inclusive, multi-dimensional account has been widely recognised [72,80,68,41]. Such an account would assist our understanding of creativity, highlighting areas of common ground and avoiding the pitfalls of disciplinary bias [34,67].

Identifying what informs our intuitive notions about creativity can guide us towards a more rigorous and comprehensive understanding. The aim of this work is to examine the nature of creativity and to identify within it a set of components representing key dimensions, as recognised across a combination of different viewpoints. This paper presents a novel, empirical approach focusing on what is revealed about our understanding of creativity's meaning and attributes through the words we use to discuss and debate the nature of creativity. Analysis of this language provides a sound basis for constructing a sufficiently detailed and comprehensive account of the concept [45,90].

In the present work, statistical language processing techniques are used to identify words significantly associated with creativity in a corpus of academic papers on the topic. A corpus of academic papers 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 of papers 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* were identified, each representing a key theme or aspect of creativity. The set of components yields information about the nature of creativity, based on what is collectively emphasised in discussions about the concept.

Within the field of computational creativity, the problem of automatic evaluation remains a significant issue: 'the Achilles' heel of AI research on creativity' [7]. The Semantic Web has emerged as a way to address the troublesome but important issue [7] of articulating values, concepts and information using an open and *machine-understandable* vocabulary. Encoding the creativity components in an OWL ontology has enabled this *ontology of creativity* to be made publicly available to the wider research community as a resource in the Semantic Web, under the permanent URL `http://purl.org/creativity/ontology`.

The remainder of this section considers the need for a definition of creativity and reviews previous attempts to provide one. The representation of subjective, ambiguous, loosely structured concepts is also considered, both from the perspective of current Semantic Web research and from a broader conceptual perspective. Section 2 provides details of the methodology used to identify components of creativity from an analysis of language data. In section 3 the results of the analysis are presented in terms of fourteen key components or dimensions of creativity. Section 4 evaluates the ontology in terms of how the components satisfy the need for a shared, encompassing, comprehensive understanding of creativity and a vocabulary to discuss aspects of creativity that is understandable both by people and by machines. Final conclusions are drawn in section 4.3.

1.1. A question of definition

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' [82, p. 43].

Many other authors have expressed similar difficulties in defining creativity and capturing a suitable definition in words [72,78,41]. Currently, no one definition has been adopted as standard, either in everyday use (as evidenced by the dictionary definitions considered in section 1.1.2 below), for research purposes [72,79,68] or even in legal practice [12,50], where one would expect a term critical to legal judgements – as creativity surely is [18,40] – to be precisely and unambiguously defined.

The difficulty of capturing an adequate definition of creativity in words should not discourage us from the attempt, however [72,68,41]. Even if other researchers have been swayed away from this task [78,83, for example], care must be taken that we 'do not throw out the baby with the bath water just because the water is cloudy' [72, p. 310]. Some of the issues concerning the problem of defining creativity have been debated for decades. The following sections consider the need for an explicit and comprehensive definition of creativity and the difficulties researchers encounter in providing one.

1.1.1. The need for a definition of creativity

Plucker makes the case that the lack of a standard definition 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' [68, 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' [68, 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' [68, 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' [72, p. 310].

In their review of research into human creativity, Hennessey and Amabile ask a significant 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?' [34, p. 570]

Two answers to this question are offered by Hennessey and Amabile, both indicating the desirability of a deeper understanding of the concept of creativity: to gain a better understanding of creativity and to learn how to boost people's creativity, both advocated as highly positive.

Creativity can and should be studied and measured scientifically, but the lack of a standard definition causes problems for measurement [41]. Plucker et al. make recommendations on best practice to follow in creativity research, based on their own survey of 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.' [68, p.92]

In short, we need to specify and justify the standards that we use to judge creativity. A more objective and specified definition of creativity enables researchers to make a worthwhile contribution [80,68,41].

1.1.2. Existing definitions

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 purposes of research the utility of such definitions is severely restricted by their format and brevity, providing 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

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

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.

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' [77, 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)' [78, p. 3]

> 'Creativity is the ability to come up with ideas or artefacts that are *new*, *surprising and valuable*' [8, 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*' [68, p. 90]

> 'Creativity: the generation of products or ideas that are both novel and appropriate' [34, 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' [72, 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.

1.1.3. 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? 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 [29,81] focus on *problem solving* and *divergent thinking* as key attributes of a creative person. In contrast, computational creativity research [64,89,65,73, for example] places emphasis on the *novelty* and *value* of creative products. Whilst there is some consensus across academic fields, 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 [8,14] or embodied and situated in an interactive environment [54,76]. Another example is whether creativity is domain-independent [66], or dependent on domain-specific context [2], or (as both Plucker and Baer have concluded) a combination of both [67,3].

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 on creative *genius* through the study of examples of people with exceptional creative achievements [69,32, for example] emphasis has shifted towards the study of *everyday* creativity, of which genius is a special case: the concept that everyone can be creative to at least some degree [88,11].

Similarly, researchers distinguish between *little-c* and *Big-C* creativity, or *psychological/P-creativity* and *historical/H-creativity* [8]), adjusting their focus accordingly to make their research more manageable. This is particularly the case in computational creativity, where the task of giving the computer general human knowledge and experiences 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 discussions thus far illuminate how creativity is a complex, multi-faceted concept that requires a broad and inclusive treatment. The *Four Ps* framework [72, 77,57,62] ensures we pay attention to four key aspects of creativity:

Person: 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 [85,69,32], tests of people's creativity [26,30,81] or tests based on creative artefact generation [1,73] are useful only within a limited sphere.

1.2. Reflections on the semantics of subjective concepts

Creativity can be seen as an essentially contested concept [22]: it is subjective, abstract and can be interpreted in a variety of acceptable ways, such that a fixed 'proper general use' is elusive [22, p. 167]. Gallie defines essentially contested concepts through several features, such as being internally complex in nature, but amenable to being broken down into identifiable constituent elements of varying relative importance, dependent on a number of factors such as context and individual preference. Though there is consensus on the concepts' meaning in very general terms, exact interpretations differ. There is not a single agreed instantiation of these concepts 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' [22, 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): what Wittgenstein refers to as 'family resemblances' [90]:

> [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.' [90, 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', but relationships between different manifestations reveal various shared characteristics emerging similar to Wittgenstein's 'family resemblances' in language. We need to identify what those family resemblances are. So to understand creativity, we can investigate what resemblances exist across different instantiations of creativity.

Wittgenstein [90] 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' [90, Part 1, Paragraph 5]. To understand the use of a word, one must have to know background information and context. Wittgenstein gives the example where a chess piece is introduced to someone as a 'king' [90, 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 set 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.

Waismann, a contemporary of Wittgenstein, has reflected on the impact of 'open texture' on language, where certain word or phrases simply cannot be completely defined for all possible scenarios:

> 'Every description stretches, as it were, into a horizon of open possibilities: however far I go, I shall always carry this horizon with me.' [84, p. 122]

Crucially for our concerns, the implications of such ambiguity mean that we are unlikely to reach the point where we have a full, complete and static definition of creativity. This causes problems for representing concepts such as creativity in a Semantic Web ontology, however it is an area where cognitive science research can considerably support and enhance progress in Semantic Web research. Cognitive linguistics research advocates that the meaning of a word is dependent on the context it is used in [17]. In particular, the premise exists in cognitive linguistics that the study of language helps reveal how people think [45,46]. 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 computational linguistics contributions [61,70,42,43].

The key principle emerging across these present discussions is that the meaning of words like creativity can best be understood by seeing how they are used by identifying different aspects that collectively contribute to the meaning of the concept of creativity. This principle guides us practically in how to approach the representation of creativity on the Semantic Web.

1.3. Subjectivity and ambiguity on the semantic web

Ambiguity in semantic interpretation of terms is not an area that Semantic Web research shies away from. Ontologies are widely used to pinpoint and define what *things* mean in a machine-readable way, for greater clarity and precision in using various vocabularies and identifiers. Such research does not concentrate merely on those areas of interest which are simple to define and objectively clear to interpret; for example SKOS [55] provides vocabulary to organise and taxonomise knowledge, whether explicit or tacit, while FRBR-oo [15] defines various terms to do with bibliographical records and the works and ideas expressed within them, on various conceptual levels beyond merely the physical instantiation of a book or other written document. Research efforts on upper ontologies such as Dolce [23], SUMO [59] and Cyc [71] have facilitated the objective, disambiguated expression of higher-level, cross-domain, more general concepts and abstract knowledge.

What is not often tackled within Semantic Web research are the types of topic discussed in section 1.2, where ambiguity is unlikely to be resolved to a single agreed definition and meanings are dynamic, changing over time and contextual interpretations, rather than being fixed to a static and unchanging prescribed definition. To date, there is a lack of suitable methodologies applied in Semantic Web research to enable or assist the building of ontological representations of such topics. Existing ontology construction methodologies tend to concentrate on how to extract and represent data objectively rather than retain subjectivity in the data [20,60,31]; the focus is on how best to construct ontologies in an appropriate manner rather than how to tackle *moving-target* issues of dynamic, perhaps unresolvable, semantics.

There is of course much work still to be done in representing semantics on the Semantic Web; much information with fixed semantics is still to be represented in a machine readable form. With the growth of the Semantic Web, this ongoing collation and representation of knowledge is work-in-progress. As discussed in section 1.2, advances in other research disciplines can assist progress in areas where there is less agreement generally over semantics. The representation of these areas and concepts on the Semantic Web is troublesome, particularly in a framework which is geared towards resolving semantic ambiguity to static, pre-assigned meanings. More established disciplines in cognitive science and related areas of research would not claim to have solved all such issues and challenges; however Semantic Web researchers can learn from advances that have been made in these disciplines. An approach informed by cross-disciplinary contributions helps Semantic Web research progress towards a broader and more complex coverage of data and their semantics.

1.4. Towards a clearer understanding of creativity

Returning to the concept under investigation in this work, the need for a clearer, multi-perspective understanding of creativity is evident (section 1.1) but remains to be addressed. There is a large quantity of material contributory to a satisfactory definition of creativity and a number of key contributions have been discussed during this section 1. What now needs to be done is to marshal this assortment of material and to unify different perspectives where possible, to avoid the disciplinary 'blinkers' or compartmentalisation that is often seen in creativity research [34]. In approaching the semantic representation of subjective and/or ambiguous concepts, some guidance is offered through philosophical reflections on the meaning of such concepts and from cognitive linguistics and computational linguistics research, as well as existing approaches to ontology learning from texts [27]. To date, however, the Semantic Web research community has largely focused on the representation of objective concepts and resources that are relatively clearly defined. Cognitive science methodologies can be harnessed to enable the representation of more semantically ambiguous concepts such as creativity, which is explored in this present work as an illustrative use case.

The approach taken for constructing this *ontology* of creativity is to collate knowledge about a concept (in this case, creativity) through an empirical study and analysis of the language used to talk about creativity. Following from the observations in section 1.2, a *confluence approach* to creativity is adopted [78,52,36]. The confluence approach works on the principle that creativity results from several components converging and examines what these components are. Taking this approach in conjunction with the application of computational linguistics methodologies and statistical analysis allows a wider disciplinary spectrum of perspectives on creativity to be captured than has previously been attempted, by breaking down the whole into smaller and more tractable constituent parts identified through a broad cross-disciplinary examination of creativity research.

2. Identifying the key components of creativity

This section describes the steps taken to discover key components of creativity. Natural language processing tools and statistical analysis methods are employed in order to identify words that appear to be highly associated with discussions 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 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.

2.1. Corpus data

A representative sample of academic papers discussing the nature of creativity was assembled. This *creativity corpus* consisted of 30 papers examining creativity from a variety of academic stand-points, ranging from psychological studies of creativity 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 Figure 2.1. 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 very recent years and which have consequently not yet accrued many citations, selection was based on subjective judgement of influence.

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 cre-

Year	Paper title	Author(s)	Reference
1950	Creativity	Guilford	*Am Psychol 5 (9), 444-454*
1960	Blind variation and selective retentions in creative thought as in other knowledge processes	Campbell	*Psychol Rev 67 (6), 380-400*
1962	The associative basis of the creative process	Mednick	*Psychol Rev 69 (3), 220-232*
1970	Identification of creativity: The individual	Dellas and Gaier	*Psychol Bull 73 (1), 55-73*
1970	Identification of potentially creative persons from the Adjective Check List	Domino	*J Consult Clin Psychol 35, 48-51*
1979	A creative personality scale for the Adjective Check List	Gough	*J Pers Soc Psychol 37 (8), 1398-1405*
1980	Primary process thinking and creativity	Suler	*Psychol Bull 88 (1), 144-165*
1983	The social psychology of creativity: A componential conceptualization	Amabile	*J Pers Soc Psychol 45 (2), 357-376*
1987	Creativity, divergent thinking, and openness to experience	McCrae and Ingraham	*J Pers Soc Psychol 52 (6), 1258-1265*
1988	Assessing everyday creativity: Characteristics of the Lifetime Creativity Scales and validation with three large samples	Richards, Kinney, Benet & Merzel	*J Pers Soc Psychol 54 (3), 476-485*
1988	Creativity syndrome: Integration, application, and innovation	Mumford and Gustafson	*Psychol Bull 103 (1), 27-43*
1988	Motivation and creativity: Toward a synthesis of structural and energistic approaches to cognition	Csikszentmihalyi	*New Ideas in Psychol 6 (2), 159-176*
1988	The creative mind: Toward an evolutionary theory of discovery and innovation	Findlay and Lumsden	*J Soc Biol Struct 11 (1), 3-55*
1992	The psychoeconomic approach to creativity	Rubenson and Runco	*New Ideas in Psychol 10 (2), 131-147*
1994	Precis of The Creative Mind	Boden	*Behav Brain Sci 17 (3), 519-570*
1995	Cognition and creativity	Runco and Chand	*Educ Psychol Rev 7 (3), 243-267*
1996	A theory of individual creative action in multiple social domains	Ford	*Acad Manage Rev 21 (4), 1112-1142*
1996	Creativity and the five-factor model	King, Walker and Broyles	*J Res Pers 30 (2), 189-203*
1996	Creativity, emergence and evolution in design	Gero	*Knowl Based Syst 9 (7), 435-448*
2000	Creativity: Cognitive, personal, developmental, and social aspects	Simonton	*Am Psychol 55 (1), 151-158*
2001	The effect of input knowledge on creativity	Colton, Pease and Ritchie	*Proc ICCBR workshop (Creative Sys)*
2002	Aspects of a cognitive theory of creativity in musical composition	Pearce and Wiggins	*Proc ECAI workshop (Creative Sys)*
2004	The cognitive neuroscience of creativity	Dietrich	*Psychon Bull Rev 11 (6), 1011-1026*
2004	Why isn't creativity more important to educational psychologists	Plucker, Beghetto and Dow	*Educ Psychol 39 (2), 83-96*
2006	A preliminary framework for description, analysis and comparison	Wiggins	*Knowl Based Syst 19 (7), 449-458*
2006	Can we trust creativity tests: a review of the Torrance Tests of Creative Thinking	Kim	*Creativity Res J 18 (1), 3-14*
2006	The transformational creativity hypothesis	Ritchie	*New Gen Com 24 (3), 241-266*
2007	Intuition, insight, imagination and creativity	Duch	*IEEE Comput Intell Mag 2 (3), 40-52*
2007	Some empirical criteria for attributing creativity to a computer program	Ritchie	*Minds Mach 17 (1), 67-99*
2009	Creativity map: Toward the next generation of theories of creativity	Ivcevic	*Psychol Aesth Creat Arts 3(1), 17-21*

Fig. 1. The 30 papers that make up the Creativity Corpus

ativity. 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 are persistent 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.

> **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[2] published in the same year, that did not contain any words with the prefix *creat* (i.e. *creativity*, *creative*, *creation*, and so on).

The non-creativity corpus is twice the size of the creativity corpus (≈ 700,000 words and ≈ 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.

[2]As categorised by the literature database *Scopus* `http://www.scopus.com/`)

- 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 [24] or an individual researcher's output [35].
- Availability of provenance data, such as who wrote the paper and for what audience (from the disciplinary classification of the journal).

2.2. *Natural language processing*

The corpus data was processed using the RASP natural language processing toolkit [10] in order to perform *lemmatisation* and *part-of-speech* tagging. Lemmatisation permits inflectional variants of a given word to be identified with a common root 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 morphological variants as an instance of the same word, rather than as separate and distinct lexical items. Lemmatisation software enables us to do this by mapping such variants to their common root 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*).

Each word was also 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 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 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*). Such words have little independent semantic content and are therefore 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. A similar list of words and frequencies was generated from the non-creativity data.

2.3. *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). The log-likelihood ratio is a measure of how well observed data fits a model or expected distribution [16,42,70,61]. 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 [16,42,61]. This is a particular issue when analysing smaller corpora, such as those used in the present work (see section 2.1). 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 [70]. Given two corpora (in our case, the creativity corpus and the non-creativity corpus) the log-likelihood score for a given word is calculated as shown in equation (1) below:

$$LL = 2 \sum_{i \in \{1,2\}} O_i \ln(\frac{O_i}{E_i}) \quad (1)$$

where O_i is the observed frequency of the word in corpus i and E_i is its expected frequency in corpus i. The expected frequency E_i is given by:

$$E_i = \frac{N_i \times (O_1 + O_2)}{N_1 + N_2} \quad (2)$$

where N_i denotes the total number of words in corpus i (i.e. the sum of the frequencies of all words drawn from corpus i).

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 creativity words may be either positively or negatively connoted. In practice this is not a problem as the significance of a given word is in its semantic connection to creativity, not its sentiment. Sentiment can be taken into account as part of the later manual examination of the data used to identify components of creativity described in section 2.4.

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.

Word (& POS tag)	LLR	Word (& POS tag)	LLR
thinking (N)	834.55	process (N)	612.05
innovation (N)	546.20	idea (N)	475.74
program (N)	474.41	domain (N)	436.58
cognitive (J)	393.79	divergent (J)	355.11
openness (N)	328.57	discovery (N)	327.38
primary (J)	326.65	originality (N)	315.60
criterion (N)	312.61	intelligence (N)	309.31
ability (N)	299.27	knowledge (N)	290.48
create (V)	280.06	experiment (N)	253.32
plan (N)	246.29	agent (N)	246.24

Table 1

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.

2.4. 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 available corpora, the resulting set of 694 creativity words is sufficiently rich for this purpose, but is somewhat too large to work with in its present state. In previous related work [37] an attempt was made to identify key components by 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 were addressed here, at least in part, by an intermediate stage of clustering all the words automatically according to a statistical measure of *distributional similarity* [48]. The more manageable collection of clusters could then be examined further to identify key components or dimensions of creativity.

The creativity words were clustered according to meaning using Lin's statistical measure of distributional similarity [48]. The intuition underlying distributional measures of similarity derives from the *distributional hypothesis* due to Harris [33]. 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 [44].

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 [47], which had previously been processed using the RASP toolkit [10] 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.[3] A possible problem with obtaining word similarity data using a corpus such as 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 and the size of the BNC corpus increased the chances of a comprehensive coverage of the general senses of each word.

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 [51, for example]. The present work adopts an information-theoretic measure due to Lin [48], which has been widely used in language processing applications as a means of automatically discovering semantic relationships between words. It has been shown to perform particularly well against other similarity measures as a means of identifying near-synonyms [86,53]. 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 Figure 2.

[3]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 [87].

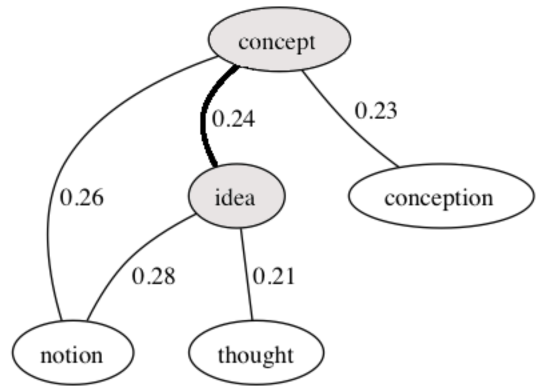

Fig. 2. Graph representation of the similarity of the nouns *concept* and *idea* and closely semantically related words. Each word i a node in the graph, linked together by a weighted edge representing the similarity of the two words (maximum similarity strength is 1.0).

Graphical representations of the similarity data like that shown in Figure 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* [5] 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.

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* [28]. 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.

Candidate components identified through clustering and inspection were further considered in terms of the

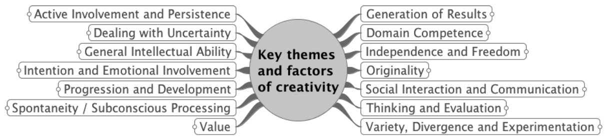

Fig. 3. The fourteen key components of creativity identified through an analysis of the word clusters

Four Ps of creativity [72,57,49,41] described earlier in section 1.1.3. 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 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.

3. Results: an ontology of creativity

3.1. 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 Figure 3 and 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 no claim is made here that the fourteen components constitutes a necessary and sufficient definition of 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 [67] 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 [13,38].

The following set of components is therefore presented as a collection of dimensions – attributes, abilities and behaviours, etc. – which contribute to an overall improvement in the understanding of creativity that provides greater clarity and information both to a human audience and in machine-readable form for future Linked Data links and applications on the Semantic Web. This approach is informed by debates in cognitive linguistics, computational linguistics and through philosophical underpinnings (section 1.2). The components should be treated as *building blocks* for creativity that may be arranged in different ways and with different emphases to suit different definitional purposes. Each component is presented here with a brief explanation; 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.

3.2. Implementing an ontology of creativity

The fourteen components provide a fuller and clearer account of the constituent parts of the concept of creativity. These components were then expressed in a machine-readable form and connected to other data sources within the Semantic Web, so that creativity becomes defined in terms of concepts that already exist on the Semantic Web. To achieve this, SKOS (Simple Knowledge Organisation System) [55] was used as a basis for constructing an OWL ontology representation of creativity and its components, in conjunction with links to the data in WordNet [19], a large lexical database of English recently made available as a Semantic Web ontology [74], in which words are grouped by sense and interlinked by lexical and conceptual relations.

The SKOS ontology incorporates three main classes: *skos:Concept* (anything we may want to record information about), *skos:ConceptScheme* (a set that collectively defines a skos:Concept) and *skos:Collection* (a collection of semantically-related information). An instance of *skos:ConceptScheme* was created as *CreativityComponents*, to represent the set of components that defines the *skos:Concept* of *Creativity*. Each component is represented as an individual *skos:Concept*. The resulting encoding can be visualised as the graph in Figure 4. The graph has also been published as an OWL/XML file at:

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

under a Open Data Commons Public Domain Dedication and Licence (PDDL) [56]. The *skos: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.

The semantics of the creativity ontology can be enhanced and supplemented through links to related ontologies and data available in the Semantic Web. The implementation of the WordNet lexical resource [19] as an ontology within the Semantic Web [74] is of particular use. Although WordNet's definition of the word

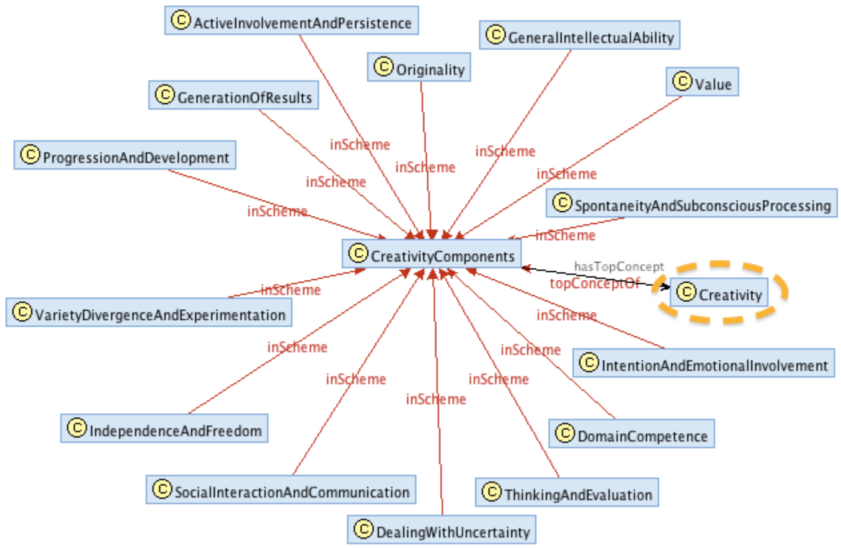

Fig. 4. The ontology of Creativity, in graph form.

'creativity' is brief ('the ability to create'), its utility is in how a lexical string (e.g. "creativity") is linked to various concepts associated with that string, such as its sense, hyponyms, type, 'gloss' (brief definition) and other related concepts.

Each component is representative of a cluster(s) of keywords from the original set of 694 creativity words. Each component has therefore been linked back to the appropriate keywords from the clusters that originally determined that component, using URIs from [74]. Following Linked Data principles, the components are hence linked across the Semantic Web through the WordNet ontology. This also provides associated semantic information on each component, by linking each component to the clusters of words that they represent. An additional step was to link `http://purl.org/creativity/ontology#Creativity` to the representation of the concept in WordNet. In this way, machines (and people) can see the relationship between WordNet's general (but brief) account of creativity and the more detailed ontological analysis in the creativity ontology.

4. Critical evaluation and discussion

As already noted above, the need for a clearer, multi-perspective and broadly applicable account of creativity is evident: this point has been repeatedly underlined by creativity researchers [72,80,68,41,34, for example]. As Hennessey and Amabile point out [34], an accurate and comprehensive account assists our understanding of creativity and further research. It also helps smooth out individual differences of interpretation, highlighting common ground and transcending discipline or domain bias [67]. The novel, language-driven approach to definition adopted in the present work draws together multiple perspectives across disciplinary divides and a confluence approach yields a broader understanding of the concept in the form of an ontology of creativity.

There are a number of ways to evaluate Semantic Web ontologies [9]. For the purposes of the creativity ontology presented in this paper, evaluation based on matching any 'golden standards' [9] is not so appropriate given the lack of any such standards or baselines to adopt, as discussed above in section 1 and elsewhere [72,79,78,68,41,39,38]. It is hoped that publication of the creativity ontology can now facilitate future applications in description and evaluation of creativity, for application-based evaluation of the ontology, perhaps incorporating other data sources to enable data-driven evaluation. In areas of knowledge representation such as that tackled in this paper, it is not appropriate to test the ontology definitions against 'a set of predefined criteria, standards, requirements, etc.' [9] as these sets do not necessarily exist, but the spirit of such recommendations can be followed by seeing how the ontology matches against human interpretations of the same concepts. One can also evaluate how the ontology data can be linked to other data sources [6]. Hence, as a reflection of the cross-disciplinary nature of this work, evaluation of the creativity ontology is two-fold, concerning both the cognitive accuracy of this representation of creativity and evaluation of the creativity data publication as an output of Semantic Web research.

4.1. Evaluation of the components as a representation of creativity

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 [38]. It is clear that a rigorous and comparative evaluation process needs clear standards to use as guidelines or benchmarks [82,41]. As part of an investigation into the nature of creativity evaluation, an evaluation study was conducted asking people to evaluate the creativity of computational musical improvisation systems [38]. A striking observation was that a number of the 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 for:

1. establishing what it means for something to be deemed creative; and
2. identifying appropriate evaluation standards that replicate typical human opinion on how creative something is or in comparing two or more creative systems.

To evaluate whether the creativity ontology presented in this paper addresses both of these requirements, the set of components has been used for the evaluation of the creativity of computational systems that improvise music [38]. The fourteen building blocks each became individual evaluation criteria. For each criterion, each system was rated by judges according to how well it met that criterion; the numeric rating was then weighted according to how important each criterion was found to be for musical improvisation creativity. Qualitative data was also collected from the judges' comments.

The results and feedback obtained in this evaluation gave an informed comparison as to which systems were more creative, and in what ways. It also found that for further improvements on the creativity of these systems as musical improvisers, greatest gains can be made in all four systems by improving performance in Social Interaction and Communication, Intention and Emotional Involvement and Domain Competence, i.e. the components found to be most important for musical improvisation creativity.

To evaluate the ontology components and compare this model of musical improvisation creativity against other models and against human intuition, the results obtained above were compared with those from applying other evaluation models and with the results of an opinion survey carried out across 111 people, who were asked how creative they thought each system was. All these evaluation approaches agreed on the most and least creative systems, but 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. In the opinion survey, participants reported difficulties in evaluating the systems' creativity. In particular, several people wanted a definition of creativity to refer to in evaluation rather than relying on their own intuitive understanding. The model offered in this paper gave the most detailed feedback, but required most information to be collected.

These evaluative results are being considered by the authors of the evaluated systems, in terms of how accurately the evaluations capture the creativity of their systems, as they perceive it, and how useful the feedback is for learning about and developing their system's creativity. Feedback so far has shown that authors find the model of musical improvisation in this paper gives the most detailed and constructive information and formative feedback on their system's creativity, with high accuracy overall except in some minor details.

The components of creativity have also been applied in an evaluation scenario where information and time was limited for evaluation; hence the components were used to simulate the forming of first impressions and snapshot judgements of the creativeness of a given computational creativity system [39]. The resulting component-based evaluation yielded detailed information about creative strengths and weaknesses, highlighting those components where a system performs strongly. Crucially, the evaluation feedback also highlighted areas where a given system performed poorly; such formative feedback is particularly useful during the development of these systems (a scenario where ongoing evaluation of progress ideally needs to be both timely and time-efficient), as insight is provided on where future development effort is best spent. Further details of this evaluation are reported in [39].

4.2. *Creativity data evaluation*

The distributed nature of Semantic Web research means that the enormous task of defining concepts in a machine-readable form is divided across the research field, rather than being the sole responsibility of one particular research group. This work practice acts as a form of peer review, as ontologies are developed, critiqued, and ultimately judged by the extent to which they are adopted and re-used as points of reference by other researchers.

The goal of the Semantic Web is to create a Web of Data [4]. The fourteen components in the creativity ontology can be considered as data about creativity published on the Semantic Web and linked to other related data sources. Berners-Lee's 'Linked Data principles' [4,6] can therefore be used to critically examine the creativity ontology in its published form on the Web of Data:

1. **'Use URIs as names for things'**
   As described in section 3.2, each 'thing' in the creativity ontology is given a URI based around the permanent URL
   `http://purl.org/creativity/ontology`, e.g.:
   `http://purl.org/creativity/ontology#Value`.
2. **'Use HTTP URIs so that people can look up those names'**
   The purl.org URIs provided for the creativity ontology are HTTP URIs.
3. **'When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)'**
   Information provided upon looking up a creativity ontology URI includes a text definition, label for that item, links to related parts of the WordNet dataset and information on how the item referenced by that URI fits into the SKOS concept hierarchy. This is provided by using the OWL standard, though, as is appropriate for an ontology, rather than a SPARQL endpoint querying a dataset stored in a RDF triplestore.
4. **'Include links to other URIs, so that they can discover more things'**
   Links are included towards relevant parts of WordNet.

4.3. *Conclusions and directions for future work*

This paper has described the development of an ontology of creativity using corpus-based, language processing techniques and its publication as machine-readable, Linked Data in the Semantic Web. The resulting ontology provides a multi-perspective analysis of creativity in terms of a set of fourteen key components and has application to the study and evaluation of computational creativity. Weaving the ontology into the Semantic Web has implications for future work on modelling subjective concepts and also feeds back into cognitive science research, suggesting some interesting approaches towards the computational modelling and automated evaluation of creativity using machine-understandable representations.

Rather than adding to the already large collection of existing definitions (section 1.1), this approach extracts common themes underlying the nature of creativity that transcend discipline or domain bias. It is observed from cognitive linguistics research and conceptual reflections on semantics (section 1.2) that words used in discussions of the nature of creativity are strongly linked to our interpretation of the meaning of creativity and its key attributes. Using techniques from corpus analysis and natural language processing (as described in section 2), key components or dimensions of creativity have been identified. The results of this novel empirical analysis, presented in section 3.1, inform the development of an ontology of creativity comprising a set of fourteen components (section 3.2). Each component makes a distinct contribution to the overall meaning of the concept. As creativity manifests itself in different ways across different domains [67], components can be varied in importance according to the requirements of a particular domain. Collectively, the components can be thought of as a set of building blocks that make creativity more tractable to study and evaluate.

The components have been applied in comparative analysis and evaluation of the creativity of computational creativity systems [38,39] (as described in section 4). The evaluative findings yielded detailed information regarding the creative strengths of each creative system evaluated, for evidence as to how that system could be considered creative. The results also highlighted weaker areas of the system, to inform future research developing each system's creativity.

Publishing the ontology in the Semantic Web ensures that it is freely available to the research community. This has a number of implications. First, it may be freely referred to, extended or amended. Refinement is clearly possible, for example in providing more fine-grained analysis of the components or in explicitly articulating the semantic relationships between them (rather than implicitly representing relationships through shared links to WordNet resources). Second, it facilitates the development of creativity-aware applications to support manual evaluation of creativity based on the components. It also represents a step towards the development of methods of automated evaluation. One intriguing possibility is to further exploit language processing techniques to provide automated evaluation by proxy based on textual reviews or descriptions of system performance. This is analogous to the way that sentiment analysis techniques are now used to automatically evaluate attitude and opinion based on reviews of products or services [63].

The motivation for this work is the need for a shared, comprehensive, multi-perspective account of creativity. Such an account can be of great value to researchers investigating the nature of creativity and in particular to those concerned with the evaluation of creative practice. If expressed in a machine-readable form, this account could contribute to what has been described as automated evaluation, 'the Achilles' heel of AI research on creativity' [7]; hence representing creativity on the Semantic Web is a contribution which can be taken advantage of by researchers studying the cognitive modelling and evaluation of creativity. In a reciprocal manner, the construction of the creativity ontology illustrates how adoption of cognitive science methods enables the modelling and representation of highly subjective, semantically contestable concepts on the Semantic Web which are unlikely to resolve to fixed, static meaning(s) (discussed in sections 1.3 and 1.4). It is hence argued that this cross-disciplinary work exemplifies a practical and applicable approach for researchers to approach the representation of these semantically dynamic, subjective and ambiguous concepts on the Semantic Web.

References

[1] Amabile, T. M. (1996). *Creativity in context*. Westview Press, Boulder, Colorado.

[2] Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal*, 11(2):173–177.

[3] Baer, J. (2010). Is creativity domain-specific? In Kaufman, J. C. and Sternberg, R. J., editors, *The Cambridge Handbook of Creativity*, chapter 17, pages 321–341. Cambridge University Press, New York, NY.

[4] Berners-Lee, T. (2009). Linked data. $http://www.w3.org/DesignIssues/LinkedData.html/$, last accessed January 2013.

[5] Biemann, C. (2006). 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*, pages 73–80, Morristown, NJ. Association for Computational Linguistics.

[6] Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22.

[7] Boden, M. A. (1999). Introduction [summary of Boden's keynote address to AISB'99]. In *AISB Quarterly - Special issue on AISB99: Creativity in the arts and sciences*, volume 102, page 11.

[8] Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Routledge, London, UK, 2nd edition.

[9] Brank, J., Grobelnik, M., and Mladenić, D. (2005). A survey of ontology evaluation techniques. In *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)*.

[10] Briscoe, E., Carroll, J., and Watson, R. (2006). The second release of the RASP system. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, Sydney, Australia.

[11] Bryan-Kinns, N. (2009). Everyday creativity. In *Proceedings of the 7th ACM conference on Creativity and Cognition*, Berkeley, California.

[12] Clifford, R. D. (2004). Random numbers, chaos theory, and cogitation: A search for the minimal creativity standard in copyright law. *Denver University Law Review*, 82(2):259–299.

[13] Colton, S. (2008). Creativity versus the perception of creativity in computational systems. In *Proceedings of AAAI Symposium on Creative Systems*, pages 14–20.

[14] Dietrich, A. and Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin*, 136(5):822–848.

[15] Doerr, M. and LeBoeuf, P. (2007). Modelling intellectual processes: The frbr - crm harmonization. In Thanos, C., Borri, F., and Candela, L., editors, *Digital Libraries: Research and Development*, volume 4877 of *Lecture Notes in Computer Science*, pages 114–123. Springer, Berlin / Heidelberg.

[16] Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

[17] Evans, V. and Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh University Press, Edinburgh, UK.

[18] Feist (1991). Feist Publications, Inc. v. Rural Telephone Service Co. *499 US 340*, 111 S. Ct 1282, 113 L. Ed. 2d 358(Supreme Court).

[19] Fellbaum, C., editor (1998). *WordNet: An electronic lexical database*. The MIT press, Cambridge, MA.

[20] Fernández-López, M., Gómez-Pérez, A., and Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In *Proceedings of AAAI'97*, Providence, Rhode Island. American Asociation for Artificial Intelligence.

[21] FOAF Project (2010). The friend of a friend (foaf) project. $http://www.foaf-project.org/$, last accessed January 2013.

[22] Gallie, W. B. (1956). Essentially contested concepts. *Proceedings of the Aristotelian Society*, 56:167–198.

[23] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). Sweetening ontologies with dolce. In

Knowledge engineering and knowledge management: Ontologies and the semantic Web, pages 223–233. Springer.

[24] Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(60):471–479.

[25] GeoNames (2010). Geonames ontology. $http://www.geonames.org/ontology/$, last accessed January 2013.

[26] Goldman, R. J. (1964). The Minnesota tests of creative thinking. *Educational Research*, 7(1):3–14.

[27] Gómez-Pérez, A. and Manzano-Macho, D. (2004). An overview of methods and tools for ontology learning from texts. *The knowledge engineering review*, 19(03):187–212.

[28] GraphViz (1988). GraphViz - Graph Visualization Software. $http://www.graphviz.org/$, last accessed January 2013.

[29] Guilford, J. P. (1950). Creativity. *American Psychologist*, 5:444–454.

[30] Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill, New York, NY.

[31] Haase, P., Lewen, H., Studer, R., Tran, D., Erdmann, M., dâĂŹAquin, M., and Motta, E. (2008). The neon ontology engineering toolkit. In *Proceedings of WWW'08*, Beijing, China.

[32] Hadamard, J. (1945). *An Essay on the Psychology of Invention in the Mathematical Field*. Princeton University Press, Princeton, NJ.

[33] Harris, Z. (1968). *Mathematical Structures of Language*. Wiley, New York.

[34] Hennessey, B. A. and Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, 61:569–598.

[35] Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16569–16572.

[36] Ivcevic, Z. (2009). Creativity map: Toward the next generation of theories of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 3(1):17–21.

[37] Jordanous, A. (2010). Defining creativity: Finding keywords for creativity using corpus linguistics techniques. In *Proceedings of the International Conference on Computational Creativity*, pages 278–287, Lisbon, Portugal.

[38] Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, 4(3):246–279.

[39] Jordanous, A. (forthcoming). *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application (submitted)*. PhD thesis, University of Sussex, Brighton, UK.

[40] Karjala, D. S. (2008). Copyright and creativity. *UCLA Entertainment Law Review*, 15:169–201.

[41] Kaufman, J. C. (2009). *Creativity 101*. The Psych 101 series. Springer, New York.

[42] Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics*, 6(1):97–133.

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

[44] Kilgarriff, A. and Yallop, C. (2000). What's in a thesaurus. In *Proceedings of the Second Conference on Language Resources and Evaluation (LREC-00)*, pages 1371–1379, Athens.

[45] Lakoff, G. (1987). *Women, Fire and Dangerous things: What Categories reveal about the mind*. University of Chicago Press, Chicago, IL.

[46] Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press, Chicago, IL.

[47] Leech, G. (1992). 100 million words of English: the British National Corpus (BNC). *Language Research*, 28(1):1–13.

[48] Lin, D. (1998). An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304, Madison, WI.

[49] MacKinnon, D. W. (1970). Creativity: a multi-faceted phenomenon. In Roslansky, J. D., editor, *Creativity: A Discussion at the Nobel Conference*, pages 17–32. North-Holland Publishing Company, Amsterdam, The Netherlands.

[50] Mandel, G. N. (2011). To promote the creative process: Intellectual property law and the psychology of creativity. *Notre Dame Law Review (online)*, 86(1).

[51] Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

[52] Mayer, R. E. (1999). Fifty years of creativity research. In Sternberg, R. J., editor, *Handbook of Creativity*, chapter 22, pages 449–460. Cambridge University Press, Cambridge, UK.

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

[54] McCormack, J. (2007). Creative ecosystems. In *Proceedings of the 4th International Joint Workshop on Computational Creativity*, pages 129–136, London, UK.

[55] Miles, A. and Bechhofer, S. (2009). SKOS simple knowledge organization system reference. $http://www.w3.org/TR/skos-reference/$, last accessed January 2013.

[56] Miller, P., Styles, R., and Heath, T. (2008). Open data commons, a license for open data. In *Proceedings of the WWW2008 Workshop on Linked Data on the Web*, Beijing, China.

[57] Mooney, R. L. (1963). A conceptual model for integrating four approaches to the identification of creative talent. In Taylor, C. W. and Barron, F., editors, *Scientific Creativity: Its Recognition and Development*, chapter 27, pages 331–340. John Wiley & Sons, New York.

[58] Music Ontology (2012). Music ontology. $http://purl.org/ontology/mo/$, last accessed January 2013.

[59] Niles, I. and Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems*, volume 2001, pages 2–9. ACM.

[60] Noy, N., McGuinness, D., et al. (2001). Ontology development 101: A guide to creating your first ontology. Technical Report Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford.

[61] Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press, Edinburgh, UK.

[62] Odena, O. and Welch, G. (2009). A generative model of teachers' thinking on musical creativity. *Psychology of Music*, 37(4):416–442.

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

[64] Pease, A., Winterstein, D., and Colton, S. (2001). Evaluating machine creativity. In *Proceedings of Workshop Program of ICCBR-Creative Systems: Approaches to Creativity in AI and Cognitive Science*, pages 129–137.

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

[66] Plucker, J. A. (1998). Beware of simple conclusions: The case for content generality of creativity. *Creativity Research Journal*, 11(2):179–182.

[67] Plucker, J. A. and Beghetto, R. A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction doesn't matter. In Sternberg, R. J., Grigorenko, E. L., and Singer, J. L., editors, *Creativity: From Potential to Realization*, chapter 9, pages 153–167. American Psychological Association, Washington, DC.

[68] Plucker, J. A., Beghetto, R. A., and Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, 39(2):83–96.

[69] Poincaré, H. (1929). Mathematical creation. In *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method.*, volume Science and Method [Original French version published 1908, Authorized translation by George Bruce Halsted], chapter III of Book I. Science and the Scientist, pages 383–394. The Science Press, New York.

[70] Rayson, P. and Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of ACL Workshop on Comparing Corpora*, Hong Kong.

[71] Reed, S. L. and Lenat, D. B. (2002). Mapping ontologies into Cyc. In *Proceedings of AAAI'02 workshop on Ontologies and the Semantic Web*, Edmonton, Canada.

[72] Rhodes, M. (1961). An analysis of creativity. *Phi Delta Kappan*, 42(7):305–310.

[73] Ritchie, G. (2007). Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17:67–99.

[74] RKBExplorer (2012). wordnet.rkbexplorer.com. $http://wordnet.rkbexplorer.com/$, last accessed January 2013.

[75] SAWS (2012). Sharing ancient wisdoms ontology for wisdom literature manuscripts. $http://purl.org/saws/ontology/$, last accessed January 2013.

[76] Sosa, R., Gero, J., and Jennings, K. (2009). Growing and destroying the worth of ideas. In *Proceedings of the 7th ACM Creativity and Cognition conference*, pages 295–304, Berkeley, California.

[77] Stein, M. I. (1963). A transactional approach to creativity. In Taylor, C. W. and Barron, F., editors, *Scientific Creativity: Its Recognition and Development*, chapter 18, pages 217–227. John Wiley & Sons, New York.

[78] Sternberg, R. J. and Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In Sternberg, R. J., editor, *Handbook of Creativity*, chapter 1, pages 3–15. Cambridge University Press, Cambridge, UK.

[79] Taylor, C. W. (1988). Various approaches to and definitions of creativity. In Sternberg, R. J., editor, *The Nature of Creativity*, chapter 4, pages 99–121. Cambridge University Press, Cambridge, UK.

[80] Torrance, E. P. (1967). Scientific views of creativity and factors affecting its growth. In Kagan, J., editor, *Creativity and Learning*, pages 73–91. Beacon Press, Boston.

[81] Torrance, E. P. (1974). *Torrance Tests of Creative Thinking*. Scholastic Testing Service, Bensenville, IL.

[82] Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In Sternberg, R. J., editor, *The Nature of Creativity*, chapter 2, pages 43–75. Cambridge University Press, Cambridge, UK.

[83] Veale, T., Gervás, P., and Pease, A. (2006). Understanding creativity: A computational perspective. *New Generation Computing*, 24(3):203–207.

[84] Waismann, F. (1968). Verifiability. In *Essays on Logic and Language*, pages 117–144. Basil Blackwell, Oxford, UK.

[85] Wallas, G. (1926). *The Art of Thought*. Jonathan Cape, London, UK, 1st edition.

[86] Weeds, J. and Weir, D. (2003). Finding and evaluating nearest neighbours. In *Proceedings of the 2nd Conference of Corpus Linguistics*, Lancaster, UK.

[87] Weeds, J. E. (2003). *Measures and Applications of Lexical Distributional Similarity*. PhD thesis, Informatics, University of Sussex, Brighton, UK.

[88] Weisberg, R. W. (1988). Problem solving and creativity. In Sternberg, R. J., editor, *The Nature of Creativity*. Cambridge University Press, Cambridge, UK.

[89] Wiggins, G. A. (2006). Searching for computational creativity. *New Generation Computing*, 24(3):209–222.

[90] Wittgenstein, L. (1958). *Philosophical Investigations, eds. Anscombe, G. E. M. and Rhees, R. and Von Wright, G. H.* Basil Blackwell, Oxford, UK, 2nd edition.