Kent Academic Repository
Full text document (pdf)

Citation for published version

DOI

Link to record in KAR

Document Version

UNSPECIFIED

Evaluating Evaluation: Assessing Progress in Computational Creativity Research

Anna Jordanous

School of Informatics, University of Sussex, Brighton, UK

a.k.jordanous(at)sussex.ac.uk

Abstract

Computational creativity research has produced many computational systems that are described as creative. A comprehensive literature survey reveals that although such systems are labelled as creative, there is a distinct lack of evaluation of the *creativity* of creative systems. As a research community, we should adopt a more scientific approach to evaluation of the creativity of our systems if we are to progress in understanding creativity and modelling it computationally. A methodology for creativity evaluation should accommodate different manifestations of creativity but also require a clear, definitive statement of the standards used for evaluation.

This paper proposes Evaluation Guidelines, a standard but flexible approach to evaluation of the creativity of computational systems and argues that this approach should be taken up as standard practice in computational creativity research. The approach is outlined and discussed, then illustrated through a comparative evaluation of the creativity of jazz improvisation systems.

Introduction

'[U]nless the motivations and aims of the research are stated and appropriate methodologies and assessment procedures adopted, it is hard for other researchers to appreciate the practical or theoretical significance of the work. This, in turn, hinders ... the comparison of different theories and practical applications ... [and] has encouraged the stagnation of the fields of research involved.'

(Pearce, Meredith, and Wiggins 2002)

In 2002 Pearce, Meredith, and Wiggins highlighted a 'methodological malaise' faced by those working with computational music composition systems due to lack of methodological standards for development and evaluation of these systems: causing progress in this research area to 'stagnate'. Computational creativity research is in danger of succumbing to this same malaise.

Computational creativity research crosses several disciplinary boundaries. The field is influenced by artificial intelligence, computer science, psychology and specific creative domains in which we implement systems, such as art, music, reasoning, story telling, and so forth (Colton 2008;

Widmer, Flossmann, and Grachten 2009; León and Gervás 2010; Pérez y Pérez 1999, provide a selection of examples).

Currently many implementors of creative systems follow a creative-practitioner-type approach: produce a system then present it to others, whose critical reaction determines its worth as a creative entity. A creative practitioner's primary aim, however, is to produce creative work, rather than to critically investigate creativity; in general this investigative aim is important in computational creativity research.

A comprehensive survey of the literature on computational creativity systems reveals the lack of systematic evaluation of the actual creativity of creative systems post-implementation. Although the quality of the system output is often subjected to some scientific evaluation, it is rare that the *creativity* of the creative system is evaluated post-implementation, or even critically commented upon (Peinado and Gervas 2006; Colton 2008, are examples of some notable exceptions). Creativity entails more than just the quality of the output: for example, what about novelty, or variety? Yet these systems are often described as creative systems without appropriate justification for this claim.

A critical analysis of current evaluation practice in computational creativity raises issues that highlight a need for a more methodical approach to evaluation to be adopted across the research community. This paper presents Evaluation Guidelines: an evaluative approach that is flexible enough to deal with different types of creativity yet allows practical and objective cross-comparison of different systems to measure progress. The Evaluation Guidelines are presented in Figure 1 and illustrated through a comparative evaluation of the creativity of jazz improvisation systems.

Computational creativity evaluation examined

To see how computational creativity systems are currently evaluated, 75 journal and conference papers were surveyed, with the aim of including all papers presenting a computational system that described that system as being creative.

Using the *Web of Knowledge* and *Scopus* databases, a literature search was conducted to find all journal papers presenting details of a computational creativity system. Words and phrases such as 'computational creativity', 'creative system', 'creative computation', 'system' and 'creativity' were used as search terms. This set of papers was supplemented with papers from journal special issues on computational

Table 1: Summary of evaluation of the 75 creative systems surveyed

Paper makes at least a mention of evaluation	77%
Paper gives details what evaluation has been done	55%
Paper contains section(s) on Evaluation	51%
Paper states evaluation criteria	69%
Main aim of evaluation: Creativity	35%
Main aim of evaluation: Quality/Accuracy/Other	43%
Mention of creativity evaluation methodology	27%
Application of creativity evaluation methodology	24%
System compared to other systems	15%
System compared to systems by other researchers	11%
Systems evaluated by independent judges	33 %

creativity (the majority of which had already been identified in the search). Reflecting the current balance of conference/workshop publications to journal publications in computational creativity research, papers from recent Computational Creativity research events were also surveyed.

Table 1 outlines the results of this survey[1]. The key finding of this survey is that evaluation of computational creativity is **not** being performed in a systematic or standard way.

Out of 75 computational systems presented as being creative systems, the creativity of a third of these systems was not even discussed when presented to an academic audience in paper format. Half the papers did not contain a section on evaluation. Only a third of systems presented as creative were actually evaluated on how creative they are.

Less than a quarter of systems made any practical use of existing creativity evaluation methodologies. Of the 18 papers that applied creativity evaluation methodologies to evaluate their system's creativity, no one methodology emerged as standard across the community. Colton's creative tripod framework (Colton 2008) was used most often (6 uses), with 4 papers using Ritchie's empirical criteria (Ritchie 2007). No other methodology was used by more than one paper.

Occurrences of evaluation being done by people outside the system implementation team were rare, as were any examples of direct comparison between systems, to see if the presented system outperforms existing systems and represents any real research progress in the field.

Why is creativity evaluation not standard practice?

By no means does this paper mean to suggest that computational creativity researchers do not wish to follow scientific practice. On the contrary, in personal communications many have expressed interest in how to evaluate creative systems, with some suggestions offered over the last decade (Ritchie 2007; Colton 2008; Pease, Winterstein, and Colton 2001).

A culture is however developing in computational creativity research where it is becoming acceptable not to evaluate the creativity of a creative system in a methodical manner.

[1] Space limitations unfortunately prevent all details being reported here; my thesis contains full survey results (Jordanous forthcoming)

To a certain extent this follows common practice of creative practitioners: to produce work then exhibit it to an audience whose reaction (both immediate and longer term) asserts the value of the work, instead of performing retrospective comparative analysis of the creativity of the work. A lack of methodical evaluation can however have a negative effect on research progress (Pearce, Meredith, and Wiggins 2002).

Evaluation standards are not easy to define. It is difficult to evaluate creativity and even more difficult to describe how we evaluate creativity, in human creativity as well as in computational creativity. In fact, even the very definition of creativity is problematic (Plucker, Beghetto, and Dow 2004). It is hard to identify what 'being creative' entails, so there are no benchmarks or ground truths to measure against.

What do we gain from scientific evaluation?

Scientific evaluation is important for computational creativity research, allowing us to compare and contrast progress. Ignoring this evaluation stage deprives us of valuable analytical information about what our creative systems achieve, especially in comparison to other systems.

Existing evaluation frameworks

Ritchie proposes empirical criteria to assess the creativity of a system based on rating the system's products for how typical of the intended genre they are and for the value of the products (Ritchie 2007). Pease, Winterstein, and Colton describe various tests of a creative system's output, input and creative process (Pease, Winterstein, and Colton 2001). Colton offers a *creative tripod* framework to qualitatively evaluate creativity (Colton 2008).

Despite these methods being available, no method has been adopted as standard evaluative practice by the research community. Colton's approach has been the most adopted by authors in the few years it has been available so far (being used to evaluate 6 surveyed systems). It is most usually used to describe why a given system should be considered creative, rather than for any comparison between systems.

As well as providing a way to evaluate the creativity of a computational system, a key function of a creativity evaluation methodology is if it enables comparison of systems against other systems, through the level of creativity demonstrated by each system. In practice, Ritchie's approach is the most frequently adopted quantitative comparison method, being applied to evaluate 4 surveyed systems.

Ritchie's proposals acknowledge several theoretical issues but are relatively impractical to use in evaluation. Several implementation decisions are left open, such as how to obtain typicality and value ratings for system products, or how to choose weights and parameter values in the criteria. Ritchie argues this allows freedom in defining creativity in the relevant domain but offers no guidelines or examples.

One other issue is how Ritchie incorporates measures of novelty (a key aspect of creativity) into the criteria. Novelty exists in more ways than whether an artefact replicates a member of the system's *inspiring set*: the artefacts that guided the construction of the system, or the inspirational material used by the system during the creative process. The criteria do not account for how surprising a product is, or

new ways of producing the end product, or how a product deviates from previous examples (Pease, Winterstein, and Colton 2001; Peinado and Gervas 2006). Also, the inspiring set may not be available for analysis, or the system may not use an inspiring set to generate new products.

The set of tests offered by Pease, Winterstein, and Colton (2001) has seen little application (perhaps due to its densely packed presentation of the test formulae). This paper has often been cited, though, and offers a considered analysis on how to evaluate computational creativity. Pease, Winterstein, and Colton admit that their choices of assessment methods are 'somewhat arbitrary' and should be treated as initial suggestions, in the hope of prompting further discussion and suggestions along similar lines. As of the time of writing, this hope has not been realised, either by the authors or by others. Of the authors of (Pease, Winterstein, and Colton 2001), only Colton makes subsequent recommendations for creativity evaluation, but these are unrelated to those in Pease, Winterstein, and Colton (2001), which is not even cited in Colton (2008).

Although not without flaws, the frameworks mentioned above and other discussions of evaluation do offer useful material for our purposes, such as the way in which the concept of creativity is broken down into constituent components and the suggestion of practical tests to carry out in evaluation. The approach to evaluation suggested in this paper aims to complement and combine the useful parts of what has been suggested so far in previous frameworks.

A reductionist approach to defining creativity

A prevalent definition of computational creativity is:

'The study and support, through computational means and methods, of behaviour exhibited by natural and artificial systems, which would be deemed creative if exhibited by humans' (Wiggins 2006)

Whilst this definition is intuitive for us to understand, it reveals little about what creativity actually is. Understanding creativity is a key aim of much computational creativity research, e.g. (Widmer, Flossmann, and Grachten 2009).

A more practical approach for detailed evaluation is taken here: that creativity is multi-dimensional, with many factors contributing to the creativity of a creative system (Pease, Winterstein, and Colton 2001; Plucker, Beghetto, and Dow 2004; Ritchie 2007; Colton 2008; Jordanous 2010a; Jennings 2010). This breaks down the concept of creativity to something more manageable and tangible, as opposed to an overarching, impenetrable concept of 'creativity'.

The need for a standard evaluation approach

A flexible approach to evaluation in this field of research is necessary. By its very nature, creativity manifests itself in a variety of forms, with different creative domains prioritising aspects of creativity differently. For the same reason, though, some standardisation is necessary to avoid the concept of creativity being interpreted too liberally, where any system could be argued to be creative depending on how creativity is defined. This approach requires that the standards used to judge creativity are stated and open to discussion.

This paper proposes a standard evaluative approach and demonstrates its application in a case study evaluating the creativity of various jazz improvisation systems. The aim of this approach is to encourage a more scientific approach to computational creativity evaluation, allowing us to identify in what areas we are achieving *creative* results and what areas we should focus more research attention on.

Standardising our approach to evaluation

Evaluation Guidelines for Computational Creativity

1. **Identify key components of creativity that your system needs if it is 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 conversely, 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. **Implement tests that evaluate your creative system under the standards stated in step 2.**

Figure 1: Proposed standard for creative systems evaluation

The intention of this approach This approach aims to examine the creativity of a creative system more systematically; to pinpoint why and in what ways a system can justifiably be said to be creative. The point is to understand to a greater level of detail exactly why a system can be described as creative. The Evaluation Guidelines approach enables us to investigate in what ways a system is being creative and how research is progressing in this area, using an informed, multi-faceted approach that suits the nature of creativity.

The Evaluation Guidelines allow comparison between a creative system and other similar systems, by using the same evaluation standards. A clear statement of evaluation criteria makes the evaluation process more transparent and makes the evaluation criteria available to other researchers, avoiding unnecessary duplication of effort.

There is a time-specific element here; a creative system is evaluated according to standards at that point in time, where a creative domain is at a certain state, viewed by society in a certain context. These standards may change over time. If similar systems have previously been presented to similar audiences at similar times, however, then the evaluation standards can be reused. Hence detailed comparisons can be made using each standard, to identify areas of progress.

What this approach is not This is not an attempt to offer a single, all-encompassing definition of creativity, nor a unit of measurement for creativity where one system may score

x %. The Evaluation Guidelines are not intended as a measurement system that finds the most creative system, or gives a single summative rating for the creativity system (though people may choose to use and adopt the approach for these purposes if it is relevant in their domain). Such a scenario is usually impractical for creativity, both human and computational. There is little value in giving a definitive rating of computational creativity, especially as we would be unlikely to encounter such a rating for human creativity.

Nor is this an attempt to dissuade researchers from attempting to implement creative systems, or to put obstacles in the way of such researchers such that they are forced to target other goals and justifications for their research rather than the pursuit of making computers creative. It is of course reasonable for computational creativity researchers to aim their work towards better understanding creativity, rather than to implement computational systems that are themselves creative. For example the pursuit of making the YQX music performance system creative (Widmer, Flossmann, and Grachten 2009) is 'abandoned' in favour of exploring human creativity via their research. However for those researchers whose intention is to implement a computer system which is creative, the approach outlined in this paper offers a methodological tool to assist progress.

Incorporating previous evaluation frameworks Depending on how creativity is defined by the researcher(s), previous evaluation frameworks (Ritchie 2007; Colton 2008; Pease, Winterstein, and Colton 2001, and other discussions) may be accommodated if appropriate for the standards by which the system is being evaluated. For example if skill, appreciation and imagination are identified as some key components of creativity for a creative system, it would be appropriate to use the creative tripod (Colton 2008).

The Evaluation Guidelines let the evaluator choose the most appropriate existing evaluation suggestions, without being tied into a fixed definition of creativity that may not apply fully in the domain they work in.

At this point no recommendations are made on what tests to include (though this paper later investigates this issue in the context of jazz improvisation systems). What is emphasised here is that for scientific evaluation we must clearly justify claims for the success or otherwise of research achievements. This approach affords such clarity.

Why not just ask humans how creative our systems are? As computational creativity is often defined as the creativity exhibited by a computational system (Wiggins 2006), experiments can be run with human judges to evaluate the creativity of a system. There is definitely a place for soliciting human opinion in creativity evaluation, not least as a simple way to consider the system's creativity in terms of those creative aspects which are overly complex to define empirically, or which are most sensitive to time and current societal context. The process of running adequate evaluation experiments with human participants, though, takes up a good deal of time and effort. Human opinion is variable; what one person finds creative, another may not (León and Gervás 2010; Jennings 2010). Therefore large numbers of participants may be required, to capture a general consensus of opinion.

In addition to the time and resources necessary to devise and run suitable evaluation experiments with large numbers of people, extra issues such as the procedure of applying for ethics permissions are introduced. There may also be some difficulty in attracting suitable participants, and a cost associated with paying participants. These issues may have adverse effects on the research process, many of which are out of our direct control to resolve. It would be useful if this outlay of research time and effort could be reduced.

There are other practical concerns which hinder us from using human judges as the sole source of evaluation of a system. Human evaluators can say whether they think something is creative but can usually give minimal insight into why it is creative. As described above, it is hard to define why something is creative; this is a tacit judgement rather than one we can easily voice. It is useful to have a more informed idea of what makes a system creative, to understand both why a system is creative and what needs to be worked on to make the system more creative.

Here one must acknowledge a common problem in computational creativity research: human reticence to accept the concept of computers being creative. On the other hand, researchers keen to embrace computational creativity may be positively influenced towards assigning a computational system more credit for creativity than it perhaps deserves. Hence our ability to evaluate creative systems objectively can be significantly affected once we know (or suspect) we are evaluating a computer rather than a human.

Implementing the Evaluation Guidelines

To illustrate how the Evaluation Guidelines approach works in practice, the approach has been applied to compare and contrast the creativity of four jazz improvisation systems:

- Voyager (Lewis 2000)
- GenJam (Biles 2007)
- EarlyBird (Hodgson 2006)
- My own jazz improvisation system (Jordanous 2010b)

Step 1a: Domain-independent aspects of creativity

To identify common components of creativity that transcend individual domains and that are applicable in all interpretations of creativity, one can look at what we prioritise as most important when we discuss creativity. This can be detected by analysing the language we use to discuss creativity, seeing what words are most prevalent in such discussions.

Previous work (2010a) identified 100 words that are most commonly used in academic literature on the nature of creativity, surveying papers across computational creativity, psychology and other disciplines to generalise across different disciplines. This work used the log likelihood ratio (Dunning 1993) to detect which words appear significantly more often in academic papers about creativity, compared to typical use in written English (as represented in the BNC).

Developing this work (Jordanous forthcoming), the same methodology was applied to compare a cross-disciplinary set of papers about creativity with a matched set of papers on subjects unrelated to creativity. This produced a list of
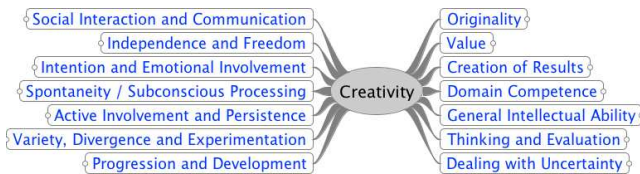
Figure 2: Key components of creativity

Figure 3: Relevance of creativity factors to improvisation

Component of creativity	weight	EarlyBird	myGA	GenJam	Voyager
Social Interaction and Communication	14.9	4.0	5.0	6.0	8.0
Domain Competence	12.5	5.0	4.0	7.0	8.0
Intention and Emotional Involvement	13.9	3.0	3.0	3.0	8.0
Active Involvement and Persistence	7.8	5.0	5.0	4.0	8.0
Variety, Divergence and Experimentation	7.1	7.0	8.0	6.0	7.0
Dealing with Uncertainty	6.4	5.0	5.0	4.0	8.0
Originality	5.8	4.0	8.0	5.0	8.0
Spontaneity and Subconscious	5.4	5.0	5.0	6.0	7.0
Independence and Freedom	5.4	4.0	6.0	3.0	8.0
Progression and Development	5.4	6.0	3.0	7.0	6.0
Thinking and Evaluation	5.1	6.0	6.0	6.0	6.0
Value	5.1	7.0	5.0	8.0	7.0
Creation of Results	3.7	8.0	7.0	7.0	9.0
General Intellect	1.4	4.0	5.0	3.0	5.0

Figure 4: Evaluating four jazz improvisation systems

words more likely to appear in the creativity literature than expected in academic papers. Grouping the results by semantic similarity, 14 key aspects or 'building blocks' of creativity are identified: see Figure 2.

Step 1b: Aspects of creativity in jazz improvisation

Berliner describes how jazz improvisers need to balance the known and unknown, working simultaneously with thought processes and subconscious emergence of ideas (Berliner 1994). Berliner examines how jazz improvisers learn from studying those who precede them, then develop that knowledge to develop a unique style.

The recent work of Louise Gibbs in jazz education equates 'creative' with 'improvisational' musicianship. She highlights invention and originality as two key components for creative improvisation (Gibbs 2010).

To identify important factors in jazz improvisational creativity, 34 participants with a range of musical experience[2] were surveyed (Jordanous forthcoming). The participants were asked to describe what creativity meant to them, in the context of musical improvisation. Their responses were grouped according to the 14 components in Figure 2.

Figure 3 summarises the participants' responses. All components were mentioned by participants to some degree. Interestingly, some components were occasionally identified as having a negative as well as positive influence. For example, over-reliance on domain competence was seen as detrimental to creativity, though domain competence was generally considered important. Of the 14 components of creativity in Figure 2, those that were identified by participants as most relevant for improvisation were:

- Social Interaction and Communication
- Domain Competence
- Intention and Emotional Involvement

Step 2: Definition of jazz improvisation creativity

Drawing upon the results from the above steps, the jazz improvisation systems were evaluated along all fourteen aspects listed in Figure 2, but with the criteria ordered so that those identified as most important were considered first, with each of the components weighted accordingly.

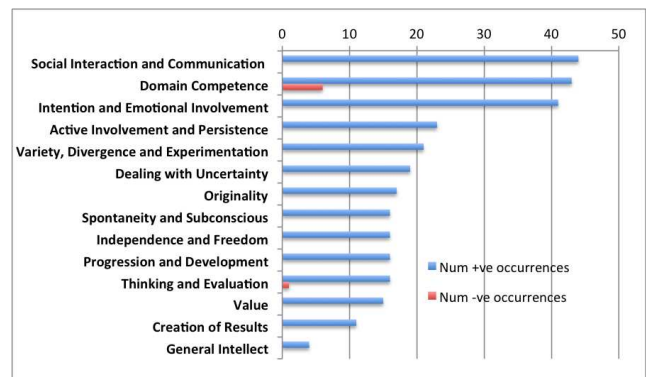[2]Musical experience: μ=20.2 yrs, σ=14.5. Improvising experience: μ=15.1 yrs, σ=14.3

Step 3: Evaluative tests for systems' creativity

Using the annotated participant data, statements were extracted to illustrate how each component is relevant to improvisation. These statements were used as test statements for each component, to analyse the four jazz improvisation systems, for example: *How is the system perceived by an audience?* (Social Communication and Interaction) *What musical knowledge does the system have?* (Domain Competence) *Does the system get some reward from doing improvisation?* (Intention and Emotional Involvement)

Each system was given a subjective rating out of 10 for each component, as represented in Figure 4. The component ratings were then weighted, so that differences in more important components were magnified, with differences in less important components reduced. This is pictured in Figure 5.

These results show that the Voyager system (Lewis 2000) can in general be considered most creative. Specifically focussing on my own system (Jordanous 2010b), while it performs well in terms of varied experimentation and in generating original results, it could be considered more creative if it was more interactive and if more musical knowledge was used during improvisation rather than random generation.

Future work and evaluation of the approach

The success of this approach can be judged by how closely it replicates creativity evaluations from human judges, so the results of applying the Evaluation Guidelines will now be
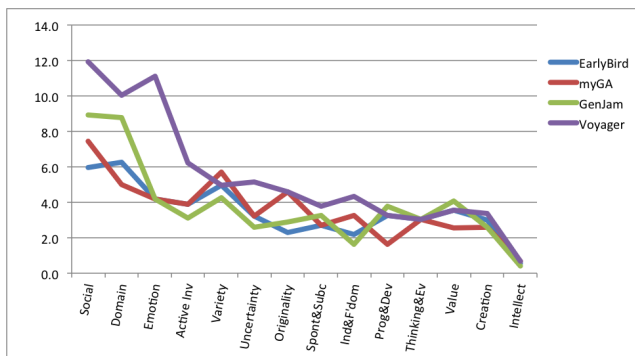
Figure 5: Weighted evaluation of the systems' creativity

compared to human evaluations of the same systems.

One reviewer of this paper commented that the Evaluation Guidelines should be applied to more domains if it is to be considered a standard evaluation methodology. I quite agree with this comment; although I am working on more applications, I hope that other researchers will consider adopting the Evaluation Guidelines to evaluate their own creative systems in other domains and share their results and observations.

Concluding remarks

A comparative, scientific evaluation of creativity is essential for progress in computational creativity. Surveying the literature on computational creativity systems, one quickly finds evidence that scientific evaluation of creativity has been neglected. While creative systems are often evaluated with regard to the quality of the output, and described as creative by the authors, in all but a third of cases the creativity of these systems is not evaluated and claims of creativity are left unverified. Often a system may be evaluated in isolation, with no reference to comparable systems.

Figure 1 presents Evaluation Guidelines, a standard but flexible approach to creativity evaluation. To demonstrate the approach, four jazz improvisation systems were comparatively evaluated to see which were more creative and, importantly, in what ways a system was more creative than another. This gave valuable information on how to improve the creativity of my own system (Jordanous 2010b).

This paper strongly advocates the adoption of the Evaluation Guidelines as standard practice in computational creativity research - to avoid computational creativity research slipping into a 'methodological malaise' (Pearce, Meredith, and Wiggins 2002).

Acknowledgements

References

Berliner, P. F. 1994. *Thinking in Jazz: The Infinite Art of Improvisation*. Chicago Studies in Ethnomusicology. Chicago, IL: The University of Chicago Press.

Biles, J. A. 2007. Improvising with genetic algorithms: *GenJam*. In Miranda, E. R., and Biles, J. A., eds., *Evolutionary Computer Music*. London, UK: Springer-Verlag. chapter 7, 137–169.

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

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

Gibbs, L. 2010. Evaluating creative (jazz) improvisation: Distinguishing invention and creativity. In Gibbs, L., ed., *Proceedings of Leeds International Jazz Conference 2010: Improvisation - jazz in the creative moment*.

Hodgson, P. 2006. Learning and the evolution of melodic complexisty in virtuoso jazz improvisation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006)*, 1506–1510.

Jennings, K. E. 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines* 1–13. Article in Press.

Jordanous, A. 2010a. Defining creativity: Finding keywords for creativity using corpus linguistics techniques. In Ventura, D.; Pease, A.; Pérez y Pérez, R.; Ritchie, G.; and Veale, T., eds., *Proceedings of the International Conference on Computational Creativity*, 278–287.

Jordanous, A. 2010b. A fitness function for creativity in jazz improvisation and beyond. In Ventura, D.; Pease, A.; Pérez y Pérez, R.; Ritchie, G.; and Veale, T., eds., *Proceedings of the International Conference on Computational Creativity*, 223–227.

Jordanous, A. forthcoming. *Evaluating Computational Creativity: A Standardised Evaluation Methodology and its Application to Case Studies*. Ph.D. Dissertation, University of Sussex, Brighton, UK.

León, C., and Gervás, P. 2010. The role of evaluation-driven rejection in the successful exploration of a conceptual space of stories. *Minds and Machines*. Article in Press.

Lewis, G. E. 2000. Too many notes: Computers, complexity and culture in Voyager. *Leonardo Music Journal* 33–39.

Pearce, M. T.; Meredith, D.; and Wiggins, G. A. 2002. Motivations and methodologies for automation of the compositional process. *Musicae Scientae* 6(2):119–147.

Pease, A.; Winterstein, D.; and Colton, S. 2001. Evaluating machine creativity. In *Proceedings of ICCBR Workshop on Approaches to Creativity*, 129–137.

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

Pérez y Pérez, R. 1999. *MEXICA: A Computer Model of Creativity in Writing*. Ph.D. Dissertation, University of Sussex, Brighton, UK.

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.

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

Widmer, G.; Flossmann, S.; and Grachten, M. 2009. YQX plays Chopin. *AI Magazine* 30(3):35–48.

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