Creativity vs quality: why the distinction matters when evaluating computational creativity systems

Anna Jordanous[1]

Abstract. The evaluation of computational creativity systems is increasingly becoming part of standard practice in computational creativity research, particularly with recent development in evaluation tools. One matter that can cause confusion, however, is in distinguishing between the concepts of creativity and quality/value. These two concepts are highly interrelated, to the point that it is difficult (and perhaps inappropriate) to define creativity without incorporating quality judgements into that definition. Several examples exist, however, where creativity evaluation has been confused with quality judgments, leading to less grounded evaluative results. Many computational creativity projects aim to produce high quality results; this is a worthy research aim. If, however, the aim of a computational creativity research project is to make as *creative* a system as possible, then a more careful approach is needed that acknowledges and understands the differences - and also the overlaps - between creativity and quality. This paper critically investigates the concepts of creativity and quality (and how they are related). It offers warning examples showing the dangers of conflating the two concepts. These are followed by practical examples of how to incorporate value judgements into the evaluation of creativity of software, to further our overall pursuit of building more creative computational systems.

1 introduction

There is a distinction to be drawn between the aim of evaluating creativity or evaluating quality (Section 3); as the survey of evaluative practice in Section 3.1 showed, these aims have become blurred to some extent.

How does one evaluate the creativity of a computer system? It can be attractive to sidestep this issue somewhat, evaluating the *quality* of a system's output rather than *creativity*.[2] As a result, though, system development progresses towards more successful output, but not necessarily more creative output. Perhaps this is what is desired? But **i**f the aim of a computational creativity research project is to make a *creative* system, then a more careful approach is needed that acknowledges and understands the differences (and the overlaps) between creativity and quality.

This paper specifically tackles the above aim, investigating at how researchers can navigate the distinction between creativity and quality, in pursuit of building more creative computational systems.

Section 2 investigates the meaning of 'quality', the meaning of 'creativity' and the ways in which these two concepts are interconnected. These investigations are carried out looking at both human and computational creativity. Section 3 tackles the question of how these concepts should be handled when performing evaluation of computational creativity systems, reflecting on relevant 'good, bad and ugly' previous evaluative practice. The paper concludes with the take-home message that while quality forms a key part of creativity, it is not a direct substitute. To evaluate creativity, we must incorporate evaluation of quality, but not treat evaluation of quality as a sufficient proxy for evaluation of creativity.

[1] University of Kent, Medway, UK, email: A.K.Jordanous@kent.ac.uk

[2] The distinction between the evaluative aims of creativity and quality is raised in Section 3.1.

2 Quality, creativity and the connections between them

2.1 Quality is...

The concept of quality, as treated in this paper, is highly related to its synonymous concepts of value as well as its near-synonyms of utility, usefulness, appropriateness, correctness, fit, relevance, and effectiveness. During this paper I will occasionally use quality and value interchangeably to represent the overarching concept of something having some worth.

Quality judgements represent the value that something has to at least one observer. As in [1], the word 'value' can be treated either as a noun or as a verb, i.e. the action of valuing; and value is not restricted to commercial or quantitative measurements, but also to cultural and qualitative assessments of value. Such judgements can be affected by societal contexts and influences, as discussed for example by Wiggins et al. in their considerations of how value is manifested in creative contexts:

> 'we treat value as a relation between an artefact, its creator and its observers and the context in which creation and observation take place.' [38, p. 2]

It is difficult to find domain-independent heuristics to follow when ascertaining the value of products. Usefulness is relative; what is considered useful in products of one domain is not necessarily reproduced in the other and may not apply equally across that individual domain. Wiggins et al [38] note in particular how Western perspectives may differ from non-Western perspectives.

To recognise the usefulness of a creative product, one might know the product's domain well enough to appreciate value, or one might have access to the opinions of people who are experts in that domain. To exemplify this point, and to begin to link quality with creativity: the greatest contributor to creativity in musical improvisation has been found to be the social communication and interaction that happens between musicians, or between performer(s) and audience during the creative process of improvising [10, 7]. Specifically for creativity, improvisers prioritise this over the quality or 'correctness' of the music produced during improvisation. In mathematical proof

derivation systems, however, accuracy (and hence quality, which is strongly related to accuracy in this domain) is vital.

Zongker's paper entitled *Chicken Chicken Chicken: Chicken Chicken* [39] demonstrates how the perception of quality in a particular domain is not always consistent across all examples of creativity in a domain. *Chicken Chicken Chicken* shows quality, in a domain that emphasises content correctness (scientific research papers), because of the extreme absence of any scientifically useful and correct content. Instead the quality of *Chicken Chicken Chicken: Chicken Chicken* comes from its value as an ironic and humorous reflection on academic publications.

2.1.1 Evaluating quality achieved by computational creativity systems

Various approaches have been used to evaluate quality; ranging from relatively simple quantitative metrics of the validity or correctness of products (as discussed below), to those more tricky evaluative scenarios when qualitative, multiple, complex or non-objective metrics are required to judge quality, as discussed for example in the [9] assessment of the cultural value of electronic musicians' creative work.

In a 2011 survey of evaluation practice in computational creativity [6, 7] (see also Section 3.1, evaluations of quality of the surveyed systems[3] were typically based on aspects of the end product(s) rather than any of the other Four Ps: process, person/producer or press (see [8]). While many examples were found of empirical measurements of value or quality, as described below, several systems were assessed for quality through user evaluations. Evaluation data was either directly provided by the user or provided indirectly through studies, such as through audience reactions and feedback at exhibitions or through qualitative tests with target users for usability and effectiveness of the system. Feedback about the appeal of systems' products and personal preferences about the products was also provided through user evaluations.

Many systems were evaluated by the correctness and validity of their products, such as calculating the percentage of material produced during runtime that can actually be used, or statistical tests for validity. Some systems were measured in terms of how interesting or novel their products were, for example seeing if the products performed at a level above a given threshold for novelty and originality in the Wundt curve function [29] or using variables representing domain-specific interest or complexity measurements.

The usefulness of a system's products could also be quantified, through the percentage of a user query which is satisfied by system output [22], or the percentage of results that are valid. Human ratings of usefulness were also used. Usefulness ratings were not all quantitative, with use of post-implementation discussions on usefulness or the interpretation of value as serving an intended purpose. Other definitions of quality were highly tuned to domain-specific metrics for value, making them less generally applicable across several types of creative system or for a more general discussion.

2.2 Creativity is...

It is difficult to define creativity without bringing quality into the definition: the concept of quality is heavily used when defining what creativity is. Psychology research has settled on a slightly controversial but now fairly commonly accepted 'standard definition of creativity' [28]:

> 'The standard definition is bipartite: Creativity requires both originality and effectiveness. ... Originality is vital for creativity but is not sufficient. ... Original things must be effective to be creative. ' [28, p. 92]

Here the word 'effectiveness' is used to represent the concept referred to in this paper as quality or value, as explained by Runco and Jaeger during the discussions in [28].

Prior to the publication of [28], the *quality* (and related concepts: value usefulness, appropriateness, relevance) and *novelty* (and related concepts: originality, newness) of creative products have often been identified as the two main aspects of creativity. Creativity was being defined in computational creativity research as 'how to create something new and useful at the same time.' [21, p. 290] Similar definitions were widely adopted e.g. in [20, 21, 27] in computational creativity, and e.g. [14, 30, 3, 23] in psychological research into creativity. Mayer [14] refers to this combination as the 'basic definition of creativity' [14, p. 450]. Table 22.1 of [14], reproduced here in Table 1, summarises the 'Two Defining Features of Creativity' [14, p. 450] as used in [30].

In a 2004[4] survey of 34 definitions of creativity used in creativity research [24], the survey found that:

> 'The most common characteristics of explicit definitions were uniqueness (n = 24) and usefulness (n = 17). Of interest, all 17 articles that included usefulness in their definition also mentioned uniqueness or novelty.' [24, p. 88]

Table 1. Mayer's summary of how novelty and value (or highly related concepts) are used to define creativity by different authors in various chapters of Robert J. Sternberg's influential *Handbook of Creativity* [14, (Table 22.1, p. 450)], in [30].

Author (Chapter)	Feature 1: Originality	Feature 2: Usefulness
Gruber & Wallace (5)	novelty	value
Martindale (7)	original	appropriate
Lumsden (8)	new	significant
Feist (13)	novel	adaptive
Lubart (16)	novel	appropriate
Boden (17)	novel	valuable
Nickerson (19)	novelty	utility

It is questionable whether the combination of novelty and value is enough to understand creativity [12]. This reductionist approach provides two tangible attributes with which to evaluate creativity. Work in computational creativity has produced countless systems that produce novel results that have value; but still the notion that computers can be creative is resisted. This undefinable part of creativity is reflected in Weiley's coining of creativity as 'novelty, value and "x" ' [32]. As argued in [11], there is much more to consider in terms of what creativity is, that the combination of novelty and value alone does not incorporate.

In dictionary definitions of creativity, the word 'quality' is one of the more frequent words used, as is 'new' (excluding common-use English words such as 'the', 'and', and so on) ([7], see also the word cloud in Figure 1). However this word cloud reveals many other

[3] Here I consider quality of a system to be treated pragmatically based on how it performs, but acknowledge that software quality in its own right, encompassing software engineering and code quality, would also be an alternative interpretation of the title of this section.

[4] This 2004 survey by [24] predates the above-mentioned work by Runco et al [28] defining their 'standard definition' of creativity.

words relating to creativity other than 'quality' and 'new'. Jordanous and Keller [11] empirically identified 14 key components of creativity through the analysis of multi-disciplinary discussions of the nature of creativity. These components do include Value, as well as Originality, but also components such as Active Involvement & Persistence, or Spontaneity & Subconscious Processing. Nonetheless, it is almost universally agreed that the concept of quality, incorporating the notions of value and usefulness, is a necessary component of creativity.

Before concluding this section, I briefly acknowledge an incidental point that connects quality and creativity in the scientific study of creativity (of which computational creativity forms a part). An interesting subjective objection to the scientific study of creativity is whether it may have a detrimental effect on our sense of the 'marvelling', 'awe and delight' of creativity:[5]

> 'Forget computers, for the moment: the conviction is that *any* scientific account of creativity would lessen it irredeemably. ... [There is a] widespread feeling that science, in general, drives out wonder. Wonder is intimately connected with creativity. All creative ideas, by definition, are valued in some way. Many make us gasp with awe and delight. ... To stop us marvelling at the creativity of Bach, Newton, or Shakespeare would be almost as bad as denying it altogether. Many people, then, regard the scientific understanding of creativity more as a threat than a promise.' [3, pp. 277-278]

3 Evaluative aims: creativity or quality?

An issue that researchers often face when evaluating their system is:

> Should systems be evaluated solely on the value and correctness of their output, or should there be some assessment of the creativity demonstrated by the system (which incorporates quality judgements on the output)?

Both are important, though the quality of output is often easier to define and test for, especially in the absence of a standard definition or creativity evaluation methodology. Particularly for computational *creativity* research, though, it is as important to consider to what extent a computational creativity system can actually be considered creative [7, 5].

It is becoming easier for computational creativity researchers to specifically target evaluation of the creativity of their systems, due to the development of evaluation tools. Creativity evaluation methods such as SPECS [7], Creative Tripod [5] or Ritchie's criteria [27] are starting to become more widely used in practice in computational creativity.

No one methodology has yet been adopted as standard, however. Historically, a 2011 survey of practice in computational creativity evaluation [6, 7] revealed issues in conflating judgements of creativity and quality during evaluation which did not follow these evaluation methods. This survey, of which the most relevant parts are reported below, investigated various questions about evaluation practice of creative systems, including these questions:[6]

- Evaluation details:
  - Is system evaluation mentioned at all in the paper?
  - Has a system evaluation been performed and described in the paper?
  - Do the authors state the aims of their evaluation and/or their evaluative criteria?
  - Is the main aim of evaluation to assess creativity (including quality of output/system) or (just) quality of output/system?
  - Brief description of evaluation done.

From the 75 surveyed creative systems in Section 3.1, only 35% of systems were evaluated according to how creative they were; the rest of the systems were evaluated solely by the quality of the system's performance. Two systems [25, 4] were described as being assessed for creativity but were actually assessed only on the accuracy of the system.

Of the 18 papers making practical use of creativity evaluation methodologies such as Ritchie's criteria [27] or Colton's Creative Tripod [5], only 10 papers used the methodologies for creativity evaluation, with the rest adapting the methodologies to evaluate the quality of their system output.

This shows some confusion about the distinction between creativity and quality; as this paper investigates, our interpretation of creativity includes reflections on quality but encapsulates more than just how correct or valuable the creative output is. A pertinent example of such confusion can be found in [31]: Ventura aimed to critically analyse creativity evaluation methodologies via a thought experiment. but actually addressed quality (or 'recognisability') evaluation only.

3.1 Survey findings

75 systems were reviewed during this comprehensive survey of computational creativity literature at the time. Results relevant to this paper are summarised in [6, 7]. Looking at the 75 surveyed systems for information relevant to this current paper:

- Of the 75 programs presented as creative systems, 26 systems (35%) were critically discussed in terms of how *creative* they were.
- 32 systems (43%) were evaluated based on the *quality* or accuracy of system performance compared to a human performing that task. This set of 32 systems includes 3 systems which were described as being assessed on how creative the systems were, but which were actually assessed by the quality of the system's performance.
- 1 paper evaluated its system in terms of knowledge gained for future research.
- The remaining 16 papers did not include evaluation of the system.

The survey also revealed interesting details from 18 papers that applied recognised creativity evaluation methodologies [5, 27, 2] or creativity models [3, 36], or that proposed new metrics to evaluate their creative systems. Of the 18 papers that applied recognised creativity evaluation methodologies:

- 10 papers used the methodologies to measure how creative their systems were.
- 6 papers adapted the chosen method to measure the quality of the systems.
- 2 papers used 'creativity' methodologies that actually measured quality.

[5] Perhaps appropriately, it is difficult to debate this point scientifically but it is worth being aware of this point, as scientific researchers studying creativity

[6] For the full list of questions included in the survey, see [7].

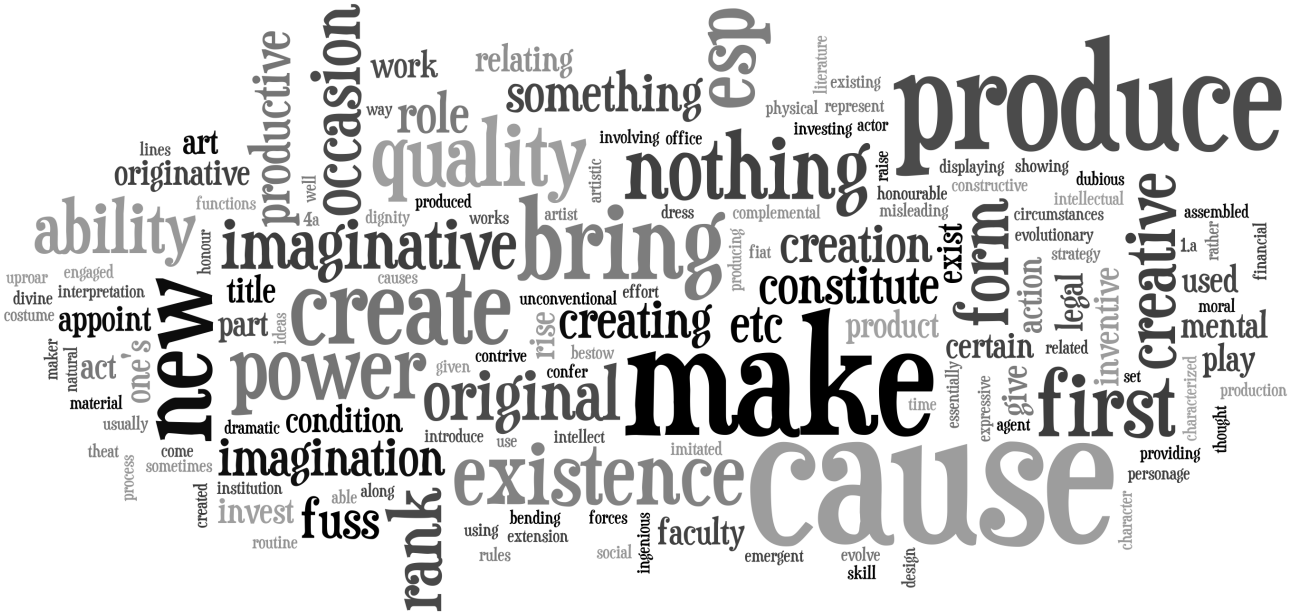

Figure 1. Words used in various dictionary definitions of creativity, as analysed in [7]. The font size of a word is relative to the frequency with which that word occurs in the collection of dictionary definitions; the larger the word, the more it appears. The word 'quality' appears fairly prominently, but does not overly dominate the diagram.

4 Using quality judgements as part of an evaluation of computational creativity

Often, as seen in the above-mentioned survey, creative systems have often been evaluated with regard to the quality of the output and this has been used to justify that system being described as creative by the authors. The discussions below first look at exemplar scenarios where such evaluation has been done due to a confusion between the two concepts of quality and creativity, then turn to discussing how quality judgements have been consciously and justifiably incorporated into evaluation of creativity.

4.1 Confusion between creativity and quality in computational creativity evaluation

The overlap between creativity and quality can sometimes cause confusion as to how to evaluate creativity; this is perhaps unsurprising given the many issues manifest in evaluating creativity of computational systems [7]. One representative example, taken from the systems surveyed above, sees a so-called creativity metric proposed which actually evaluates quality, but which is used to generate evidence justifying a system being labelled a creative system.

Collins et al. [4] employ the Wilcoxon's two-sample statistical test on their music harmonisation generator. The metric examines similarity between generated output and the system's knowledge base. Despite Collins et al. describing their test as a creativity metric, they actually measure how closely the system can replicate the test set (similar to the approach in [34]). In other words, in [4] what is actually proposed is a correctness metric rather than a creativity metric, a distinction which they briefly acknowledge as they admit a lack of conviction in describing their system as creative:

> 'This paper has presented a metric for evaluating the creativity of a music-generating system. Until further evaluation has been conducted (by human listeners rather than just by the creativity metric), we are cautious about labelling our overall system as creative.' [4, p. 9]

Another example of this confusion is clear when we ask people to evaluate the creativity of computational systems; expressions of the difficulty of this task is often acknowledged by recognition that the evaluators do not know where a creativity judgement is distinct from a value judgement. For example, in the case studies reported in [7] where several systems were evaluated on how creative they were, quotes from respondents included:

> 'I liked this one better than the other ones, but am really struggling to distinguish between "like" or "approve" and "think it's creative".'
>
> 'I kept going with my gut instinct which was basically to rate it on how much I *liked* it... but I don't think that really equates to how creative it was... but I'm not even sure a computer *can* be creative, which is why I had to just keep reverting to "like".'
>
> 'I think it depends on the definition of creativity - is it just creating something? or creating something that makes the appropriate amount of "sense", for want of a better word, for people to appreciate? I'm using the latter definition!'

There are deeper issues afoot here than mere confusion, to do with people's perception of computational creativity.[7] One evaluator in the [7] evaluation case studies explained how they struggled with applying the concept of creativity to computers when they saw creativity as 'a uniquely human thing'. Thus, they instead resorted to a conceptually easier measure of aesthetic, even though they were aware of the difference, concluding:

> 'I preferred these samples to the previous ones, but that isn't really a measure of creativity!'

[7] This thorny topic will not be explored in this current paper, but is explored to some extent in [7] as well as in [15, 13, 16].

4.2 Conscious incorporation of quality evaluation in creativity evaluation

The blurring of evaluative aims, between assessing quality and creativity, is a theme that is detectable not only in [4], but also in several computational creativity system evaluations.[8] Several evaluation tools in computational creativity, however, consciously include quality evaluation as part of a creativity evaluation for creative systems. This is in keeping with the view that quality or value is a fundamental component of creativity [28, 11].

4.2.1 Ritchie's empirical criteria for computational creativity

Graeme Ritchie proposed a set of formal empirical criteria for creativity [27]. The criteria are situated in an overall framework describing the design and implementation of a creative computational system in set-theoretic form. Ritchie advocates post-hoc analysis of artefacts generated by the system, disregarding the process by which they were created. For systems that produce abstract rather than concrete results (Ritchie gives the example of analogies), Ritchie's approach is not applicable.

The criteria collectively describe aspects of the *typicality* and *quality* of the output of the creative system (and indirectly, the novelty of the system output). Two key mappings are used to separate out the concepts of typicality and novelty:

typ - a rating of how typical the output is in the intended domain

> 'To what extent is the produced item an example of the artefact class in question?' [27, p. 73]

val - a rating of how valuable the output is

> 'To what extent is the produced item a high quality example of its genre?' [27, p. 73]

Ritchie emphasises the importance of assessing computer-generated artefacts both in terms of how typical an example they are of items in the target domain and in terms of atypicality. Further to this, an artefact may be typical of the domain but not be a good example, so the value rating is introduced to assess the quality of that artefact.

> 'If a person produces a painting which is radically different from previous work ... and which is definitely a good painting, then that will usually be deemed creative. ' [26, p. 4]

The formal definitions of the 18 criteria can be found in [27]. Here, the criteria are deliberately presented informally, with descriptors such as 'suitable' and 'high' substituted for the parameters left unspecified in [27]. It is hoped that any subsequent loss in formal semantics is balanced by a more immediate understanding of each criterion. We can see that while Ritchie allows for both typical and atypical results to be recognised by the criteria, criteria involving value judgements are always required to find high levels of value if that criterion is to be satisfied.

1. On average, the system should produce suitably typical output.
2. A decent proportion of the output should be suitably typical.
3. On average, the system should produce highly valued output.
4. A decent proportion of the output should be highly valued.
5. A decent proportion of the output should be both suitably typical and highly valued.
6. A decent proportion of the output is suitably atypical and highly valued.
7. A decent proportion of the atypical output is highly valued.
8. A decent proportion of the valuable output is suitably atypical.
9. The system can replicate many of the example artefacts that guided construction of the system (the *inspiring set*).
10. Much of the output of the system is not in the inspiring set, so is novel to the system.
11. Novel output of the system (i.e. not in the inspiring set) should be suitably typical.
12. Novel output of the system (i.e. not in the inspiring set) should be highly valued.
13. A decent proportion of the output should be suitably typical items that are novel.
14. A decent proportion of the output should be highly valued items that are novel.
15. A decent proportion of the novel output of the system should be suitably typical.
16. A decent proportion of the novel output of the system should be highly valued.
17. A decent proportion of the novel output of the system should be suitably typical and highly valued.
18. A decent proportion of the novel output of the system should be suitably atypical and highly valued.

4.2.2 Pease et al.'s tests on the input, output and process of a system

[20] proposed a combination of evaluative tests for creativity, based on:

- The input provided to the system.
- The output produced by the system.
- The process(es) employed by the computational system.

The tests for the output produced by the system are categorised by Pease et al. as either *Novelty Measures* or *Quality Measures*. The latter set of quality measures consists of:

- Quality Measures.
  - *Emotional Response Measure:* human judges evaluate to what degree an item has affected them positively or negatively; the responses are used to categorise items according to the intensity of the response.
  - *Pragmatic Measure:* using unspecified (domain-specific) 'marking criteria' [20, p. 6] to judge to what extent an item meets an aim.

The tests for the process(es) employed by the computational system are divided into two sets: tests of generative processes and tests of evaluative processes within the systems. The evaluation set of tests includes a test based around a quality judgement:

- *Evaluation of Process Measure:* comparing the quality measures from above on two comparable sets of output items. One set is produced by methods which can be transformed internally during program run-time and one by methods which cannot. The quality of the first set should exceed the quality of the second.

[8] In fact, encountering such examples during peer review duties was one of the prompts to write this paper.

As in Section 4.2.1, these tests are summarised in informal language above.[9]

4.2.3 Wiggins' framework for categorising creative systems

Wiggins proposed a framework for categorising creative systems [37] inspired by Boden's proposals on creativity [3]. Strictly speaking, this framework is for formal description and classification of different aspects of creative system, rather than evaluation of creativity, but has been used in computational creativity evaluation.

Though the framework is intended to be used to 'analyse, evaluate and compare creative systems' [36, p. 1], Wiggins carefully states that he does not contribute to the debate on creative evaluation:

> 'I am making no attempt here to discuss or assess the value of any concepts discovered: while this issue is clearly fundamentally important [citing [3, 26, 17]], it can safely be left for another time.' [37, p. 453]

The framework describes system details formally according to seven formal rule sets and functions relating to the system's conceptual space (i.e. the set of all possible items that could conceivably be output by the system). One of these rulesets, $\mathcal{E}$, is the set of rules used to evaluate items in the conceptual space. This set, as with the others, is left to be populated as the framework is applied for categorising creative systems.

Looking at Wiggins' immediate subsequent work as a guide for how to populate this set, Wiggins and colleagues have tended to focus on quality evaluation rather than creativity evaluation [34, 19, 33]. In particular [19]'s melody generation system was evaluated using a variation of Amabile's Consensual Assessment Technique (CAT) [2], intended by Amabile for evaluating the creativity demonstrated by humans in a quantitative way by expert judges. CAT was adapted slightly in [19] to assess the quality of output ('stylistic success') from their system rather than the creativity of the system itself. This decision was perhaps influenced by the authors' substantial background in musical quality evaluation e.g. [35, 17, 36, 18, 37].

5 CONCLUSION

Distinguishing between creativity and quality is a tricky task to negotiate when we are evaluating the creativity of computational creativity systems. The two concepts overlap considerably; in fact it is generally accepted that creativity cannot be defined without incorporating the concept of quality into that definition. The two concepts are however not to be confused; an evaluation of value is distinct from an evaluation of creativity.

Above, examples have been presented where such confusion in evaluation has led to less-than-solid conclusions about the results of such evaluation. We have also seen, however, that creativity evaluation can (and should) incorporate evaluation of quality as part of that overall evaluation. Quality is a necessary part of creativity; but it is not sufficient for creativity.

ACKNOWLEDGEMENTS

These thoughts have been shaped over conversations with several people over the past years, most recently a discussion point raised by Oded Ben-Tal at the 2017 Computational Simulation of Musical Creativity conference on whether creative music was the same as high quality music.

[9] For more formal descriptions of the tests, the interested reader can refer to [20].

REFERENCES

[1] Daniel Allington, Byron Dueck, and Anna Jordanous. Networks of Value in Electronic Dance Music: SoundCloud, London, and the Importance of Place, 2015.
[2] Teresa M Amabile, *Creativity in context*, Westview Press, Boulder, Colorado, 1996.
[3] Margaret A. Boden, *The creative mind: Myths and mechanisms*, Routledge, London, UK, 2nd edn., 2004.
[4] Tom Collins, Robin Laney, Alistair Willis, and Paul Garthwaite, 'Music: Patterns and Harmony Using Discovered, Polyphonic Patterns to Filter Computer-generated Music', in *Proceedings of the International Conference on Computational Creativity*, pp. 1–10, Lisbon, Portugal, (2010).
[5] Simon Colton, 'Creativity versus the Perception of Creativity in Computational Systems', in *Proceedings of AAAI Spring Symposium on Creative Intelligent Systems*, pp. 14–20, Stanford. CA, (2008). AAAI Press.
[6] Anna Jordanous, 'Evaluating Evaluation: Assessing Progress in Computational Creativity Research', in *Proceedings of the Second International Conference on Computational Creativity (ICCC-11)*, Mexico City, Mexico, (2011).
[7] Anna Jordanous, *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*, Ph.D. dissertation, University of Sussex, Brighton, UK, sep 2012.
[8] Anna Jordanous, 'Four PPPPerspectives on computational creativity in theory and in practice', *Connection Science*, **28**(2), 194–216, (2016).
[9] Anna Jordanous, Daniel Allington, and Byron Dueck, 'Measuring cultural value using social network analysis: A case study on valuing electronic musicians.', in *ICCC*, pp. 110–117, Park City, UT, (2015).
[10] Anna Jordanous and Bill Keller, 'What makes musical improvisation creative?', *Journal of Interdisciplinary Music Studies*, **6**(2), 151–175, (2012).
[11] Anna Jordanous and Bill Keller, 'Modelling creativity: identifying key components through a corpus-based approach', *PloS ONE*, **11**(10), e0162959, (2016).
[12] James C Kaufman, *Creativity 101*, The Psych 101 series, Springer, New York, 2009.
[13] Carolyn Lamb, Daniel G Brown, and Charles LA Clarke, 'Human competence in creativity evaluation.', in *ICCC*, pp. 102–109, (2015).
[14] Richard E Mayer, 'Fifty Years of Creativity Research', in *Handbook of Creativity*, ed., Robert J Sternberg, chapter 22, 449–460, Cambridge University Press, Cambridge, UK, (1999).
[15] David C Moffat and Martin Kelly, 'An investigation into people's bias against computational creativity in music composition', in *The Third Joint Workshop on Computational Creativity*, Riva del Garda, Italy, (2006).
[16] Philippe Pasquier, Adam Burnett, Nicolas Gonzalez Thomas, James B Maxwell, Arne Eigenfeldt, and Tom Loughin, 'Investigating listener bias against musical metacreativity', in *Proceedings of the Seventh International Conference on Computational Creativity*, (2016).
[17] Marcus Pearce and Geraint Wiggins, 'Towards a framework for the evaluation of machine compositions', in *Proceedings of the AISB Symposium on AI and Creativity in Arts and Science*, York, UK, (2001).
[18] Marcus T Pearce, *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*, Ph.D. dissertation, Department of Computing, City University, London, UK, 2005.
[19] Marcus T Pearce and Geraint A Wiggins. Evaluating Cognitive Models of Musical Composition, 2007.
[20] Alison Pease, Daniel Winterstein, and Simon Colton, 'Evaluating Machine Creativity', in *Proceedings of the ICCBR'01 Workshop on Creative Systems*, pp. 129–137, Vancouver, Canada, (2001).
[21] F Peinado and P Gervas, 'Evaluation of automatic generation of basic stories', *New Generation Computing*, **24**(3), 289–302, (2006).
[22] Francisco C Pereira and Amílcar Cardoso, 'Experiments with free concept generation in Divago', *Knowledge-Based Systems*, **19**(7), 459–470, (2006).
[23] Jonathan A Plucker and Ronald A Beghetto, 'Why Creativity is Domain General, Why it Looks Domain Specific, and why the Distinc-

tion Doesn't Matter', in *Creativity: From Potential to Realization*, eds., Robert J Sternberg, Elena L Grigorenko, and Jerome L Singer, chapter 9, 153–167, American Psychological Association, Washington, DC, (2004).

[24] Jonathan A Plucker, Ronald A Beghetto, and Gayle T Dow, 'Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research', *Educational Psychologist*, **39**(2), 83–96, (2004).

[25] Mark O Riedl, 'Vignette-Based Story Planning: Creativity Through Exploration and Retrieval', in *Proceedings of the 5th International Joint Workshop on Computational Creativity*, pp. 41–50, Madrid, Spain, (2008).

[26] Graeme Ritchie, 'Assessing Creativity', in *Proceedings of the AISB Symposium on AI and Creativity in Arts and Science*, pp. 3–11, York, UK, (2001).

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

[28] Mark A. Runco and Garrett J. Jaeger, 'The standard definition of creativity', *Creativity Research Journal*, **24**(1), 92–96, (2012).

[29] Rob Saunders, Petra Gemeinboeck, Adrian Lombard, Dan Bourke, and Baki Kocaballi, 'Curious Whispers: An Embodied Artificial Creative System', in *Proceedings of the International Conference on Computational Creativity*, pp. 100–109, Lisbon, Portugal, (2010).

[30] *Handbook of Creativity*, ed., Robert J Sternberg, Cambridge University Press, Cambridge, UK, 1999.

[31] D Ventura, 'A Reductio Ad Absurdum Experiment in Sufficiency for Evaluating (Computational) Creative Systems', in *Proceedings of the 5th International Joint Workshop on Computational Creativity*, pp. 11–19, Madrid, Spain, (2008).

[32] Viveka Weiley, 'Remixing realities: distributed studios for collaborative creativity', in *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, pp. 345–346, Berkeley, California, (2009). ACM.

[33] Raymond Whorley, Geraint Wiggins, Christophe Rhodes, and Marcus Pearce, 'Development of Techniques for the Computational Modelling of Harmony', in *Proceedings of the International Conference on Computational Creativity*, pp. 11–15, Lisbon, Portugal, (2010).

[34] Raymond P Whorley, Geraint A Wiggins, and Marcus T Pearce, 'Systematic Evaluation and Improvement of Statistical Models of Harmony', in *Proceedings of the 4th International Joint Workshop on Computational Creativity*, pp. 81–88, London, UK, (2007).

[35] Geraint Wiggins, Eduardo Miranda, Alan Smaill, and Mitch Harris, 'A Framework for the Evaluation of Music Representation Systems', *Computer Music Journal*, **17**(3), 31–42, (1993).

[36] Geraint A Wiggins, 'Categorising creative systems', in *IJCAI*, Acapulco, Mexico, (2003). IJCAI.

[37] Geraint A Wiggins, 'A preliminary framework for description, analysis and comparison of creative systems', *Knowledge-Based Systems*, **19**(7), 449–458, (2006).

[38] Geraint A. Wiggins, Peter Tyack, Constance Scharff, and Martin Rohmeier, 'The evolutionary roots of creativity: mechanisms and motivations', *Philosophical Transactions of the Royal Society B*, **370**, 20140099, (2015).

[39] D Zongker, 'Chicken Chicken Chicken: Chicken Chicken', *Annals of Improbable Research*, **12**(5), 16–21, (2006).