Co-creativity and perceptions of computational agents in co-creativity

Anna Jordanous
School of Computing, University of Kent
Chatham Maritime, Medway, Kent, UK
a.k.jordanous@kent.ac.uk

Abstract

How are computers typically perceived in co-creativity scenarios? And how does this affect how we evaluate computational creativity research systems that use co-creativity? Recent research within computational creativity considers how to attribute creativity to computational agents within co-creative scenarios. Human evaluation forms a key part of such attribution or evaluation of creative contribution. The use of human opinion to evaluate computational creativity, however, runs the risk of being distorted by conscious or subconscious bias. The case study in this paper shows people are significantly less confident at evaluating the creativity of a whole co-creative system involving computational and human participants, compared to the (already tricky) task of evaluating individual creative agents in isolation. To progress co-creativity research, we should combine the use of co-creative computational models with the findings of computational creativity evaluation research into what contributes to software creativity.

Introduction

'This experiment ... has plainly benefited from a lot of human intervention ... To call it "computer-generated" is misleading. "Computer-initiated" and "computer-assisted", though less grabby, are more accurate - and in their own way provide a thought-provoking novelty' (Telegraph)[1]

'[Humans] are, essentially, curating and correcting the computers' output' (What's on Stage)[2]

'... the computer generated claim starts to unravel. There's no software that can put all of these elements together and turn them into a musical. That requires a human' (Engadget UK)[3]

'I do think the fundimentals [sic] are being nudged by humans' (kevlawdrums on Twitter, Mar 3 2016)

'it's absolutely brilliant - but had a lot of human creative input too' (SemarkJ on Twitter, Feb 27 2016).

These quotes relate to the 'Beyond the Fence' musical theatre project (Colton et al. 2016), in which various computational creativity systems created parts of a musical. Comments on the project included criticism that the project's billing as 'the world's first computer-generated musical' was misleading; it was 'computer-assisted' but not computer-generated. The validity of describing the 'Beyond the Fence' project as an example of *co-creativity* between humans and computers was not explored (Jordanous in press).

In a co-creative scenario involving human and computational participants, how do we attribute and evaluate the creative contributions of different types of participants? In particular, how do we evaluate computational participants relative to human participants, or relative to the system as a whole? Although there have been many advances in computational creativity evaluation tools to date, typically these tools tend to assume we are evaluating a single piece of software, working without creative contribution from other entities. In co-creativity, there are more than one participants contributing to the creative process, but often we cannot delineate the specific contribution of each participant.

As discussed below, Bown (2015) has suggested attributing creative agency to different participants via studying the dynamics of interaction and activities, following Maher's suggestions (Maher 2012) on determining individual and collective creativity. While useful as an approach for how to analyse the contributions of different participants to creativity, Bown acknowledges his proposal is 'despite its long standing in social sciences, quite a radical approach to thinking about attributing creative agency. This view removes the privilege of the human actor' (Bown 2015, p. 21)

Is computational creativity research at a stage where we can afford to remove this 'privilege of the human actor' without consequence, or should we be more cautious in how the contributions of creative systems are attributed in co-creative scenarios? Following Colton (Colton 2008), who closely links the *perceived* creativity of a system to evaluation of creativity, we can rephrase this question:

Is there a problem of bias affecting how well we can recognise computational participants' contributions to co-creative scenarios? If so, how could/should computational creativity researchers address this problem, to make it easier to evaluate and acknowledge computational contributions to co-creativity?

[1] http://www.telegraph.co.uk/theatre/what-to-see/beyond-the-fence-arts-theatre-review-computer-says-so-so/

[2] http://www.whatsonstage.com/london-theatre/reviews/beyond-the-fence-arts-theatre_39847.html

[3] http://www.engadget.com/2016/03/02/beyond-the-fence-computer-generated-musical/

This paper investigates this two-fold question, first considering what problems we should be concerned about in attribution of creativity to computational co-creativity participants. The term co-creativity is established and different manifestations of co-creativity are explored. To understand better what might be needed for an entity to be treated as a potential co-creative participant, co-creativity scenarios between humans and computers are compared to scenarios of collaboration between human participants, and to scenarios involving creativity support tools. Broader questions are also considered, on how to acknowledge computational agents' creativity and how our confidence in attributing creativity is affected by issues of biases. An experimental case study provides data to complement these more theoretical discussions, analysing how people judge a computational system in a co-creative scenario compared to if they were under the impression that this computational system was working in isolation (i.e. not in a co-creative scenario). The paper concludes by discussing the implications for computational co-creativity research; how can we evaluate the creativity of co-creative computational agents, and justify our conclusions - particularly to those not minded towards accepting the possibility of computational creativity?

Background

In a 2005 journal special issue on human-computer creativity, Lubart[4] provides a commentary on 'How can computers be partners in the creative process' (Lubart 2005). This focus quickly moves, however, to the subtly but distinctly different in meaning: 'focus on ways that computers can contribute to people's creativity'. This de-emphasises the concept of people contributing to computers' creativity, or of people and computers being treated as comparable creative contributors. A 2010 follow-up (Burkhardt and Lubart 2010) shows no particular concessions towards the idea that computers could play a more creative role in co-creativity. To echo Ada Lovelace's words from two centuries earlier, computers are seen as having 'no pretensions whatever to originate anything. It [the Analytical Engine] can do whatever we know how to order it to perform. ... Its province is to assist us in making *available* what we are already acquainted with' (Lovelace 1843) (emphasis in original).

More criticism arose around the possibility of a computational participant to be able to contribute to a set of creative processes in more than a merely supplemental, assistive way. One comment from critics reviewing the recent 'Beyond the Fence' musical theatre project (Colton et al. 2016) was that the project's billing as 'the world's first computer-generated musical' was in fact an inaccurate claim; the project should instead have been described as 'computer-assisted', not 'computer-generated' (Jordanous in press). As discussed elsewhere (Jordanous in press), this was probably a fair criticism that leads us to another useful question: how could computational participants have made a more genuinely co-creative contribution to this project?[5]

[4]This special issue predates by 11 years Lubart's keynote at the 2016 International Conference on Computational Creativity.

[5]This specific question is tackled in Jordanous (in press).

To find evidence to argue for creative computational contributions to co-creativity, we could make some attempt to evaluate the creativity of computational agents using existing or new evaluation methods in computational creativity.[6] This is problematic, however. Firstly, existing methods are poor for evaluating individual parts of a larger system (Kantosalo, Toivanen, and Toivonen 2015; Jordanous in press), though this has recently been investigated by drawing upon evaluation methods from interaction design (Kantosalo, Toivanen, and Toivonen 2015). Secondly, and the focus of this paper: evaluation of co-creative software can be heavily affected by a deeper question around relative perceptions of creativity in co-creative scenarios between humans and computers. We want to better understand perceptions within co-creativity and make sure we fully capture what evidence is needed to justify claims that the co-creative software participant(s) make valid contributions to co-creative scenarios.

What is co-creativity? Definition

Co-creativity can be modelled in many ways (Candy and Edmonds 2002) but the basic requirement is that more than one participant collaborates actively in a creative process. In computational creativity research on co-creativity, at least one of these participants is computational. Davis (2013) treats co-creativity as a gestalt-like process, where the creative participants contribute in a way which 'the sum is greater than the parts' (rather than a 'division of labor model' where tasks are divided up and allocated to participants in the process (Davis 2013, p. 10). He sees co-creativity as a 'blend' of collaborative improvisation by human and computational agents, 'where a human and computer both make creative contributions to an artwork' (Davis 2013, p. 9). In particular, Davis focuses on a supposed 'new type of human-computer co-creativity' (Davis 2013, p. 9):

> 'that introduces a computer into this collaborative environment as an equal in the creative process ... in which the human and computer improvise in real time to generate a creative product. Here, creativity emerges through the interaction of both the human and the computer. ... The contributions of human and computer are mutually influential, which is the nature of collaboration and improvisation.' (Davis 2013, p. 10)

This is modelled computationally using an enactive model of cognition (Davis et al. 2015). Further examples of co-creativity in computational models and creative software are starting to emerge (Liapis, Yannakakis, and Togelius 2012; Yannakakis, Liapis, and Alexopoulos 2014; Magerko et al. 2014; Grace and Maher 2014; Kantosalo, Toivanen, and Toivonen 2015; Jacob and Magerko 2015).

Different types of co-creativity

In addressing the question 'How can computers be partners in the creative process' (Lubart 2005), four different roles are identified that computers can take in human-computational co-creativity:

[6]For an overview and discussion of such methods, see the chapters by Ritchie and by Jordanous in the forthcoming 'RADIANCE' Springer volume of Readings in Computational Creativity.

- 'Computer as nanny' [p. 366] (helping people to work as efficiently as possible during creative tasks, for example by minimising distractions).
- 'Computer as penpal' [p. 367] (an enabler for helping humans to record and communicate their thoughts/ideas).
- 'Computer as coach' [p. 367] ('an expert system, knowledgeable in creativity-relevant techniques', to help the user to be more creative).
- 'Computer as colleague' [p. 368] ('a real partnership, in which humans and computers work hand in hand').

Only the fourth of these roles suggests that computers could play anything other than a restricted role in the creative process; however, sadly, even the discussion of 'a real partnership' between computers and humans in the creative process is tempered with the assertion that 'Computers can probably better implement random searches than humans but humans are needed to select the best ideas and perhaps to fine hone these ideas, turning them into viable creative productions'. In other words, computers cannot be creative (according to Lubart) except as an assistant for searching and generating possibilities at random, at which point the human must take over in order for the products of such a partnership to be considered 'viable creative productions'.

What is necessary in co-creativity for an entity to be recognised as a co-creative participant? Specifically in the context of co-creativity, we can consider how to recognise the creativity of computational participants through analogy. How is creativity of individual participants recognised in co-creativity between human participants? What can we learn from creativity support tools, that are designed specifically to support human creativity? And what methods help us attribute computational contributions to co-creativity?

Comparison with how we treat human collaborators in a creative team comprised fully of human participants Davis et al. (2015) use human collaboration and improvisation as the basis for their model of human-computer co-creativity. They discuss a '*creative trajectory*, which is the shared understanding and intention to make creative contributions in a mutually negotiated and desired direction' (Davis 2013, p. 11). Co-creativity in general is therefore treated by Davis et al. as a set of meaningful interactions between creative partners, in a call-and-response type model where each partner responds to the other's communications.

Comparison with creativity support tools (CST) Davis sees human-computational co-creativity as bridging a 'gap in the CST research literature on how computer colleagues can contribute to the creative process' (Davis 2013, p. 9). Creativity support tools (Schneiderman 2007) focus on supporting human creativity and disregard computational creativity. CST research tackles questions such as: *'What tools, methodologies, and practices can support creativity of individuals in interdisciplinary teams?'* (Mamykina, Candy, and Edmonds 2002, emphasis in original). It is seen as an HCI research area, not computational creativity/AI (Davis 2013).

Recognising computational agents' creativity Maher asks the question 'Who's Being Creative?' in co-creative scenarios with collaborations between different entities (Maher 2012). In these collective creativity scenarios, Maher points out that creative responsibility can be assigned either to individuals within the co-creative scenario or to the collective of entities involved. Maher proposes measuring levels of *ideation* and *interaction* (suggestions overlapping Bown's later suggestions somewhat (Bown 2015)). Systems can be categorised and ordered along two dimensions of *ideation*, measured along axes that range from modelling to generating, and from the ability to suggest, through enhancement of ideas and to generation. *Interaction* modelling and categorisation is similarly measured in terms of the number of entities involved, human and computational.

Kantasalo et al. argue for a more complex approach: human-computer co-creativity evaluation should be conducted using mixed methods (Kantosalo, Toivanen, and Toivonen 2015), with human opinion a crucial part of analysis. To recognise creativity in an entity, of course, objective systematic evaluation is not necessarily required. Eigenfeldt et al. recognise (at least) five ways of validating the creativity of a system, by considering the perspectives of: the architect of a system; the audience who engage with creative outputs by the system; the academics engaging with publications about a system; domain experts engaging with the system as peers and critics; and the results of controlled system experiments (Eigenfeldt, Burnett, and Pasquier 2012).

Bown's commentary emphasises the need to recognise the contributions of different agents in the creative process (Bown 2015) and looks at examples of how to do this. He also reminds us that such recognition of creative contribution is necessary in the context of single computational creativity systems: 'the "islands of creativity" problem' (Bown 2015, p. 18) highlights the misconception that creativity occurs solely within the bounds of a single creative agent (human or computational). Creative agents engage with the outside world in many ways, as per the 'Press/Environment' aspect of the Four Ps of creativity (Jordanous 2016; Rhodes 1961). Creative agents are influenced by external sources, e.g. as audience or peer; and systems influence the world around them by what they are doing. This is captured in creativity models such as the DIFI domain-individual-field-interaction model (Csikszentmihalyi 1988) or Hennessey and Amabile's (2010) systems take on creativity theory.

A dynamic analysis of creative systems has been offered to analyse creative contributions from different entities (Bown 2015), which defines creativity and the production of output via (a) the interactions between agents and (b) the notion of *becoming*: i.e. how things might be continually dynamically re-created, rather than existing in a stable static form. There is development from Maher's (2012) proposals to model creative systems via their ideation abilities and interaction. In Bown's examples of how to attribute creative agency to participants, he outlines how the creative process can be broken down into dynamic activities over a particular timeline, and see how different activities involve or influence different entities. As example, he analyses the creativity of different participants in a musical performance involving two human musicians and various algorithmic performers. The collaboration is broken down into steps, helping to identify where participants influence each other. (Here the system designer is also considered a participant).

In situations where bias is carefully controlled, Bown's analysis provides detailed qualitative descriptions of creative contributions by the participants in the described scenarios. Sadly, such situations are rare in practice; the approach passes over the problems of bias (conscious or subconscious) that arise in evaluation of computational creativity when computational participants' creativity is examined. The next section investigates these problems in more detail, focusing on the context of co-creativity.

Bias and confidence in recognising creativity

To what extent do bias and differing levels of confidence influence research processes (particularly evaluation) in computational creativity? In Davis's work on co-creativity, it is interesting to contrast (1) his repeated emphasis on the equality of the partnership between human and computer in co-creativity with (2) occasional deviations away from this viewpoint in his writing, towards a view of the computer as the more subservient partner in this process, supporting human creativity without necessarily recognising any creativity on the part of the computational agent. For example (emphases added): 'The proposed system, Coco Sketch, encodes some rudimentary stylistic rules of abstract sketching and music theory to contribute *supplemental* lines and music while the user sketches.' (Davis 2013, p. 9). Or: 'Creative computers could understand and work alongside humans in a new hybrid form of human-computer co-creativity that could inspire, motivate, and perhaps event each creativity to *human users* through collaboration.' (Davis et al. 2015, p. 109) Evaluation in (Davis 2013) focuses on discussing/reporting and measuring the creativity of the human participants in a human-computer co-creative scenario, with the only evaluation of the computational participant's contribution being the use of Amabile's Consensual Assessment Technique to evaluate the creative output.

There is 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. Recognition of computational creativity in general is however affected by a number of different issues and challenges, many of which are matters of significant, long-standing and/or continuous debate both in the computational creativity community and in research on human creativity. The perception of a computer system as being creative (or potentially creative) is a different issue to the question of whether a computer system actually is creative (Colton 2008). Both of these issues are relevant in this current description of the recognition of the creativity of computational co-creativity participants, in particular the first. Creativity has been described as being 'in the eye of the beholder' (Cardoso, Veale, and Wiggins 2009, p. 17). As emphasised in the Four Ps approach to creativity (Jordanous 2016), the opinions of the audience play a key part in making, distributing and maintaining creativity judgements. Overcoming negative preconceptions about computational creativity can therefore be a necessary hurdle for computational creativity researchers to negotiate.

Practically it can be tricky to use human judges for evaluating creativity of a computational system. Human evaluators can make a tacit judgement on whether they think something is creative but may find their decision difficult to explain or justify (as investigated in Jordanous (2012b)). In a case study asking people to assess the creativity of systems (Jordanous 2012b), several participants requested a definition of creativity. This study also showed the variance in human opinion; what some found creative, others did not. While larger studies might capture consensus of opinion if it exists, consensus is not necessarily a guaranteed result (Jordanous 2012a; 2012b).

People's competence in evaluating computational creativity can be questioned (Lamb, Brown, and Clarke 2015); evaluations can be influenced by preconceived notions and beliefs. This perception has been discussed from several different perspectives e.g. (Minsky 1982; Moffat and Kelly 2006; Colton 2008; Nake 2009; Reffin-Smith 2010; Pease and Colton 2011; Jordanous 2012b). People may be reluctant to accept the concept of computers being creative, either through conscious reticence or subconscious bias (Moffat and Kelly 2006). Researchers keen to embrace computational creativity may be positively influenced to assign a computational system more credit for creativity than it perhaps deserves. In short, 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. *Perhaps we evaluate systems differently to how we evaluate people - and perhaps systems' collaboration with people distorts how we attribute creativity to that system?*

Experimental work: an evaluative case study

The above questions probe how we recognise the creativity of computational participants in co-creative scenarios, to assist us in computational creativity evaluation. It would be useful to have data to make comparative judgements of systems' creativity in a co-creative context compared to outside this context. In an ideal world, we could simply ask people directly to compare the creativity of a system in a co-creative context to its creativity when working on its own. Life is not that simple, though; it is difficult to ask directly without encountering subconscious bias or confusion over what it means for computers to be creative (Jordanous 2012b).

Instead, in this case study, the aim is to collect a rough consensus of opinion on how creative a system is: either with the participants under the impression that the system is operating independently on its own; or being informed that the system is operating in a co-creative scenario with a human user. This study investigates the two-part hypothesis:

> *i. People will vary in their evaluation of how creative a system is considered to be, depending on the evaluator's impression of whether it is operating on its own or whether it is operating in collaboration with human(s).*
> *ii. People are more confident attributing creativity to computational systems operating in isolation, than attributing creativity to computational systems that operate within a co-creative scenario.*

Such comparisons let us investigate if people consider a co-creative system differently if treated as part of a complete co-creative collaboration of human plus computational participant (instead of evaluating the computational participant in isolation from the human collaborator). The null hypothesis, hence, is that there is no significant difference in evaluation or in confidence of evaluation of a creative system on whether it is co-creative or not, and regardless of if an entire co-creative scenario is being evaluated or whether we are evaluating individual participants within that scenario.

We compare the evaluations (and participants' reported confidence in evaluations) of the same system with people under different impressions as to how much it collaborated with other people, and compare the evaluations of the system in isolation to the system as part of a bigger co-creative system of collaboration. We also compare confidence levels for evaluations of a co-creative system (either treating the 'system' (1) as all participants, human and computational, or (2) just evaluating the computational participant), against confidence in evaluating a computational system that works in isolation, with no human-computer interaction.

Method

Three groups of participants were used for this study, each with a minimum of 30 participants. Participants were asked to indicate their opinions on how creative a co-creative computer system is, and were also asked to indicate how confident they were about their answer. The participants were given a brief description of how the system works (using non-technical language and avoiding jargon). They were also provided with sample outputs of the system.

The co-creative system used in this study is the Impro-Visor (Keller 2012) musical improvisation system, which works in conjunction with human student users to generate jazz music in the style of improvised solos. Impro-Visor is intended as a tool to help people develop their ability to construct jazz solos. It learns musical grammars from corpora of jazz music and uses those as generative grammars to suggest solos which can be developed by the human user.[7]

The three groups were divided according to how they were asked to approach this task:

- Group 1: Treat Impro-Visor as part of a co-creative system, evaluated separately from the human user participating in the co-creative process
- Group 2: Treat Impro-Visor + human user, considering both as parts of a human-computer co-creative system
- Group 3: Treat Impro-Visor as a standalone creative system, *not* a co-creative system

Group 1 evaluated Impro-Visor specifically as the computational participant in a co-creative scenario, i.e. considering the software participant's creativity. Group two evaluated how creative the co-creative system was as a whole, i.e. Impro-Visor and the human user. Data for both groups 1 and 2 were collected during lectures on computational creativity evaluation, to students at undergraduate or MSc level. The exercises were completed at the start of these lectures, before the students had been given any information from the lecturer on computational creativity evaluation; the teaching purpose of the exercise was to get students to think practically about how to evaluate creative systems, and prompt them to consider to themselves the issues that arise, prior to any discussions of this in the teaching session.

The third group's data was collected slightly differently. In fact, this study was originally inspired by a previous - and aborted - data collection for an evaluative case study involving Impro-Visor, conducted under a false impression that the output samples used for the case study were indeed generated purely by the system in question working independently. Later it was discovered (personal communications with Bob Keller, 2012), that the system outputs were the result of the system and a human user in collaboration. This was unfortunate for the original study, but a serendipitous inspiration and source of data for this study. This original data on evaluation of Impro-Visor seemingly as a standalone system, not collaborating with a human user but operating autonomously, was collected via an online survey for (Jordanous 2012b) that presented participants with descriptions of four musical improvisation systems and their outputs. Participants were asked how creative they thought each system was, and to give their confidence in their decision.

Groups 1 and 2 evaluated Impro-Visor as a second task after performing a similar evaluation of a non-co-creative system Tale-Spin (Meehan 1976). The main purpose of the Tale-Spin evaluation was mostly to give the participants a practice evaluation task. Though it also provides data to compare the confidence of participants in groups 1 and 2 at evaluating a co-creative system compared to a non-co-creative system, such analysis is limited; many factors influence why we may be more confident evaluating one system than another (as shown by comments made in the study).

Of course the evaluation of one system compared to another unrelated system in a different domain does not give us any sort of complete picture as to how people treat systems in co-creative scenarios compared to non-collaborative scenarios. While the data from evaluation of Tale-Spin is interesting as some sort of yardstick, the evaluation task for Tale-Spin is essentially a practice exercise before the main exercise we are interested in collecting data for.

For the evaluative exercise, the participants were asked: "How creative do you think the *XXX* system is?" (substituting for *XXX* the system perspective they were being asked to consider, based on whether they were in group 1, 2 or 3). Ratings of creativity were collected using a five-point labelled Likert scale 0-4: [0] Not at all creative, [1] A little creative but not very, [2] Quite creative, [3] Very creative, [4] Completely creative. Then participants were asked: "How confident are you about the answer you just gave for *XXX* system's creativity?" (substituting for *XXX* the system perspective they were being asked to consider). Ratings of confidence were also collected using a five-point labelled Likert scale 0-4: [0] Very unconfident, [1] Unconfident, [2] Neutral, [3] Confident, [4] Very confident.

[7] Although Impro-Visor recently gained functionality for collaborative real-time improvisation with a user (Keller et al. 2012), this functionality was not included in the evaluation case study.

Table 1: Data on ratings of Impro-Visor creativity on the 0-4 scale, from 'Not at all creative' to 'Completely creative'

Group	1	2	3
N	34	50	120
mean	1.91	1.93	1.88
median	2.00	2.00	2.00
mode	2.00	2.00	2.00
std dev	0.723	0.874	0.913

Table 2: Data on confidence of participants in providing ratings of Impro-Visor creativity. On the 0-4 scale, this ranges from 'Very unconfident' to 'Very confident'

Group	1	2	3
N	34	50	120
mean	2.44	2.02	2.74
median	2.50	2.00	3.00
mode	3.00	3.00	3.00
sd	0.801	1.008	0.912

Results of case study

A summary of the collected data is given in Tables 1 and 2. A two-tailed t-test was used to compare whether there were significant differences in ratings in confidence between the three groups, with significance set to $p \leq= 0.05$.

The first set of tests tested the null hypothesis that there was no significant difference in ratings for Impro-Visor's creativity, across the three groups (i.e. first comparing group 1 and group 2, then comparing group 1 and group 3, then finally comparing group 2 and group 3). No significant evidence was found to disprove the null hypothesis; in other words none of the three different perspectives taken on Impro-Visor significantly affected how creative the system was perceived to be. Across each of the three groups, the Impro-Visor system was rated at between 1.88 mean (group 3) and 1.93 (group 2). This implies a rating of between 2 ('Quite creative') and 1 ('A little creative but not very').

The second set of tests tested the null hypothesis that there was no significant difference in confidence for rating Impro-Visor's creativity between pairwise combinations of the three groups. The alternative hypothesis is that a significant difference in confidence has been found.

In comparing groups 1 and 3, i.e. comparing the confidence in rating the Impro-Visor software individually as a co-creative participant and as a standalone piece of software that does not collaborate with a human, a p value of 0.06 meant that although close to being significant, there is no statistical evidence for a significant difference in confidence. In the other two pairwise combinations, however, significant evidence was found to reject the null hypothesis.

Comparing groups 2 and 3, i.e. considering the collective creativity of Impro-Visor and the human user compared to the creativity of Impro-Visor as a standalone program, a p-value of 0.00002 indicated strongly significant evidence of a difference in confidence levels. Remembering that a confidence rating of 3 represents an answer: 'Confident', and 2 represents an answer: 'Neutral', a mean of 2.74 for confidence in ratings for group 3 contrasts against a mean of 2.02 for group 2. **In other words, participants were significantly more confident at rating the creativity of Impro-Visor if they were under the false impression that it was operating on its own, without user collaboration - compared to trying to rate the collective creativity of Impro-Visor and the human user together.**

Participants were also significantly lower in confidence (p=0.038) at completing the latter task compared to the confidence levels showed by group 1 participants, who had to rate the creativity of Impro-Visor as an individual participant in this co-creativity scenario.

Using t-tests, comparisons were also made between the confidence of participants rating Impro-Visor's creativity compared to their confidence in rating Tale-Spin's creativity, for participants in groups 1 and 2 who evaluated both systems. However no significant evidence was found to show that participants felt more or less confident rating the co-creative system Impro-Visor, compared to the standalone system Tale-Spin. This was as expected, due to the many factors that influence confidence in evaluating systemsbased on differences in processes, products and domain.

Discussion

Regarding the hypothesis under investigation The data shows that while there seems to be little significant difference in the level of creativity attributed to Impro-Visor, people typically felt significantly less confident about evaluating the creativity of a co-creative system of human and computer participants, compared to evaluating the creativity of the computational participant, or if they were under the impression that the co-creative system was actually working as a standalone system with no human collaborator.

One explanation of the data may be that the participants would generally feel less confident about assigning creativity to a collaborative group of participants, compared to attributing creativity to individual participants within that creative scenario. To understand retrospectively how participants decided on their rankings, we can look at the qualitative data collected during the case study described above.

Many participants took advantage of the option to add qualitative comments to support their evaluation data. In the group 3 evaluation scenario, though many comments were given on the musicality of the system, only one participant from the 120 participants involved made a comment about their confidence in their answer:

> *'I liked this one better than the other ones, but am really struggling to distinguish between "like" or "approve" and "think it's creative".'* [User rated Impro-Visor as 3 for creativity and 3 for confidence]

In groups 1 and 2, more participants commented on their confidence in performing these evaluative tasks. For those in group 2, where participants were evaluating the collective co-creative system of software plus human, no comments at all were made about any difficulties in attributing creativity collectively to a pair of creative collaborators rather than to the individual participants involved. Of the comments that were made, the following are interesting for showing people's attitudes to the computational participant in the co-creative scenario [their ratings for creativity of Impro-Visor and their indicated confidence are given in brackets]:

> *"The system has the student to help out on creativity."* [creativity=1, confidence=3]
>
> *"These systems can't evaluate their own creativity. The main difference is the human input"* [creativity=1, confidence=2]

In group 1, where the participants were typically lower in confidence than the other two groups in their evaluations, evaluating the collective co-creative system of Impro-Visor plus its human user, similar indications were still given as to the relative creative contributions of each participant. One participant underlined the reference to the human user in their answer of how creative the system of 'computer+human' was, presumably to indicate that their attribution of a creativity rating of 2 was because of the human participant (given with a confidence rating of 2). Other interesting comments made by this group's participants included:

> *"The music sounds like it could have been made by a human alone whereas the story seemed computer generated rather than human written."* [creativity=3, confidence=3]
>
> *"Quite creative based on the human influence."* [creativity=2, confidence=1]
>
> *"human element helps with the confidence and knowing more about the workings."* [creativity=3, confidence=3]
>
> *"If I take into account the fact that a computer was involved I tend to (rather silly and unrealistically) consider the whole system less creative. Bias towards human creativity!"* [creativity=2, confidence=1]
>
> *"Impro-Visor: Don't know how much input came from the musician (pre-knowledge?) and how much came from the system?"* [creativity=2, confidence=1]

Despite these comments, the data on creativity ratings does not show that people gave *significantly* higher ratings when considering the human as a part of the co-creative system being evaluated (group 2, with a mean of 1.93 for creativity rating compared to 1.91 for group 1 and 1.88 for group 3). Those who added extra comments, however, wished to indicate their reliance on the human part of this co-creative system. Unsurprisingly, perhaps, no participants indicated the opposite sentiment that the computational element helped them feel more confident about their evaluation than if evaluating only the human participant.

Conclusions and implications for future work

How are computers typically perceived in co-creativity scenarios? The study reported here supports the literature review suggestions that participants are less confident attributing creativity in collective co-creativity scenarios including computational participants, arguably because of a reluctance to assign creative agency to computational participants.

The data from the case study in this paper shows that while there seems to be little significant difference in the level of creativity attributed to Impro-Visor, people typically felt significantly less confident about evaluating the creativity of a co-creative system of human and computer participants, compared to evaluating the creativity of the computational participant, or if they were under the impression that the co-creative system was actually working as a standalone system with no human collaborator. People typically felt 'Neutral' in their confidence of judging the collective creativity of a human-computer collaboration, compared to half way between 'Neutral' and 'Confident' for rating the system individually in this co-creative scenario, and three-quarters of the way from 'Neutral' to 'Confident' for participants' average confidence in rating Impro-Visor's creativity if they had thought the system was working autonomously.

Bown has suggested a working model for how to attribute creative agency to different participants in human-computer co-creativity, based on interactions and dynamic tracing of influence in the creative activities (Bown 2015). This builds on previous suggestions to model co-creative systems individually or collectively via measuring ideation and social interactions in the systems (Maher 2012) and mixed methods (Kantosalo, Toivanen, and Toivonen 2015). There is a certain level of naivety to these approaches - understandable given the desire to move forwards in a more objective and methodical approach for attributing creative agency in co-creativity. With these approaches, we have a decent theoretical starting point; but nonetheless further attention needs to be paid to practical issues that arise in their application.

Human evaluation should indeed be used to recognise and assess the contribution of computational participants in human-computer co-creativity (Kantosalo, Toivanen, and Toivonen 2015). What, though, needs to be done for computers to be perceived as genuine partners in a co-creative process, making a *creative* contribution? How can we demonstrate that in creative scenarios, computational software is not merely limited to the remit of creativity support tools, supporting human creativity, but can become (and exceed the requirements of) a creative 'colleague' (Lubart 2005)? Can we even aim for a point at which human creativity can be seen to support computational creativity?

These are not simple questions to answer, but they should not be passed over. With computational creativity models and software, we can, and have, explored such questions further, to advance understanding of co-creativity more generally. As discussed above, ongoing evaluative research on what makes software appear creative (or what makes software actually creative?) helps us pursue these questions. Once we better understand how to deal with inherent conscious/subconscious biases involved in human evaluation of computational creativity, our working approaches for creativity attribution become more useful as a basis for accurately recognising creative agency in co-creative software.

References

Bown, O. 2015. Attributing creative agency: Are we doing it right? In *Proceedings of the 6th International Conference on Computational Creativity*. Park City, UT: ACC.

Burkhardt, J.-M., and Lubart, T. 2010. Creativity in the age of emerging technology: Some issues and perspectives in 2010. *Creativity and Innovation management* 19(2):160–166.

Candy, L., and Edmonds, E. 2002. Modeling co-creativity in art and technology. In *Proceedings of the 4th International Conference on Creativity and Cognition*, 134–141. Loughborough, UK: ACM.

Cardoso, A.; Veale, T.; and Wiggins, G. A. 2009. Converging on the Divergent: The History (and Future) of the International Joint Workshops in Computational Creativity. *AI Magazine* 30(3):15–22.

Colton, S.; Llano, M. T.; Hepworth, R.; Charnley, J.; Gale, C. V.; Baron, A.; Pachet, F.; Roy, P.; Gervas, P.; Collins, N.; Sturm, B.; Weyde, T.; Wolff, D.; and Lloyd, J. 2016. The beyond the fence musical and computer says show documentary. In *7th International Conference on Computational Creativity (ICCC 2016)*. Paris (France): ACC.

Colton, S. 2008. Creativity versus the Perception of Creativity in Computational Systems. In *Proceedings of AAAI Symposium on Creative Systems*, 14–20. Stanford, CA: AAAI.

Csikszentmihalyi, M. 1988. Society, culture, and person: a systems view of creativity. In Sternberg, R. J., ed., *The Nature of Creativity*. Cambridge, UK: Cambridge University Press. chapter 13, 325–339.

Davis, N.; Hsiao, C.-P.; Popova, Y.; and Magerko, B. 2015. An enactive model of creativity for computational collaboration and co-creation. In Zagalo, N., and Branco, P., eds., *Creativity in the Digital Age*. London, UK: Springer. 109–133.

Davis, N. 2013. Human computer co-creativity: Blending human and computational creativity. In Smith, G., and Smith, A., eds., *Proceedings of the Doctoral Consortium of Artificial Intelligence and Interactive Digital Entertainment Conference (AIIDE) 2013*, number WS-13-23 in AAAI Technical Report, 9–12. Boston, MA: AAAI.

Eigenfeldt, A.; Burnett, A.; and Pasquier, P. 2012. Evaluating musical metacreation in a live performance context. In *International Conference on Computational Creativity*, 140.

Grace, K., and Maher, M. L. 2014. Towards computational co-creation in modding communities. In *Proceedings of AIIDE*. AAAI.

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

Jacob, M., and Magerko, B. 2015. Interaction-based authoring for scalable co-creative agents. In Toivonen, H.; Colton, S.; Cook, M.; and Ventura, D., eds., *Proceedings of the Sixth International Conference on Computational Creativity*, 236–243. Park City, UT: ACC.

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

Jordanous, A. 2012b. *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application*. Ph.D. Dissertation, University of Sussex, Brighton, UK.

Jordanous, A. 2016. Four PPPPerspectives on computational creativity in theory and in practice. *Connection Science* 28(2):194–216.

Jordanous, A. in press. Has computational creativity successfully made it 'Beyond the Fence' in musical theatre? *Connection Science*.

Kantosalo, A.; Toivanen, J. M.; and Toivonen, H. 2015. Interaction evaluation for human-computer co-creativity: A case study. In *Proceedings of the 6th International Conference on Computational Creativity*. Park City, UT: ACC.

Keller, R. M.; Toman-Yih, A.; Schofield, A.; and Merrit, Z. 2012. A creative improvisational companion based on idiomatic harmonic bricks. In Maher, M. L.; Hammond, K.; Pease, A.; Pérez y Pérez, R.; Ventura, D.; and Wiggins, G., eds., *Proceedings of the Third International Conference on Computational Creativity*, 155–159. ACC.

Keller, R. M. 2012. Continuous improvisation and trading with Impro-Visor. In Maher, M. L.; Hammmond, K.; Pease, A.; Pérez y Pérez, R.; Ventura, D.; and Wiggins, G., eds., *Proceedings of the Third International Conference on Computational Creativity*. ACC.

Lamb, C.; Brown, D. G.; and Clarke, C. 2015. Human competence in creativity evaluation. In Toivonen, H.; Colton, S.; Cook, M.; and Ventura, D., eds., *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)*, 102–109. Park City, Utah: Brigham Young University.

Liapis, A.; Yannakakis, G. N.; and Togelius, J. 2012. Co-creating game content using an adaptive model of user taste. In Maher, M. L.; Hammond, K.; Pease, A.; Pérez y Pérez, R.; Ventura, D.; and Wiggins, G., eds., *Proceedings of the Third International Conference on Computational Creativity*. ACC.

Lovelace, A. 1843. Notes on Manabrea's Sketch of the Analytical Engine Invented by Charles Babbage. In Bowden, B., ed., *Faster than thought : a symposium on digital computing machines (1953)*. London: Pitman.

Lubart, T. 2005. How can computers be partners in the creative process: Classification and commentary on the special issue. *International Journal of Human-Computer Studies* 63:365–369.

Magerko, B.; Permar, J.; Jacob, M.; Comerford, M.; and Smith, J. 2014. An overview of computational co-creative pretend play with a human. In *Proceedings of the Playful Characters worksop at the Fourteenth Annual Conference on Intelligent Virtual Agents*.

Maher, M. L. 2012. Computational and Collective Creativity: Who's Being Creative? In *Proceedings of the 3rd International Conference on Computer Creativity*. Dublin, Ireland: ACC.

Mamykina, L.; Candy, L.; and Edmonds, E. 2002. Collaborative creativity. *Communications of the ACM* 45(10):96–99.

Meehan, J. R. 1976. *The metanovel: writing stories by computer.* Ph.D. Dissertation, Yale University, New Haven, CT, USA.

Minsky, M. L. 1982. Why people think computers can't. *AI Magazine* 3(4):3.

Moffat, D. C. D. C., and Kelly, M. 2006. An investigation into people's bias against computational creativity in music composition. In *The Third Joint Workshop on Computational Creativity*.

Nake, F. 2009. Creativity in algorithmic art. In *Proceedings of the 2009 Conference on Creativity and Cognition*, 97–106.

Pease, A., and Colton, S. 2011. On impact and evaluation in Computational Creativity: A discussion of the Turing Test and an alternative proposal. In *Proceedings of the AISB'11 Convention*. York, UK: AISB.

Reffin-Smith, B. 2010. 43 Dodgy Statements on Computer Art. http://zombiepataphysics.blogspot.com/2010/03/43-dodgy-statements-on-computer-art.html, last accessed 29th July 2010.

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

Schneiderman, B. 2007. Creativity support tools: accelerating discovery and innovation. *Communications of the ACM* 50(12):20–32.

Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2014. Mixed-initiative co-creativity. In *Proceedings of the 9th Conference on the Foundations of Digital Games*.