

Has computational creativity successfully made it ‘Beyond the Fence’ in musical theatre?

Anna Jordanous

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

Abstract

Beyond the Fence is a commercial project, undertaken for a television documentary, that has produced a musical show billed as “the world’s first computer-generated musical”¹. Several computational creativity systems have been used in the production of various parts of this musical, which has been performed in London’s West End for a two week run in 2016. Having been involved in this project as an informed commentator who was not involved in creating any of the software, I consider two questions that together form the main contribution of this paper: (1) To what extent is the project successful? and (2) To what extent does this project demonstrate computational creativity? Investigations into these questions show that *Beyond the Fence* has successfully shown how existing creative software can indeed be used to create a plausible and acceptable musical. The resulting musical has been moderately well-received by most critics, though standards are raised high for computational creativity in the public eye. The project has also raised the profile of computational creativity research. Some useful lessons have also been highlighted for computational creativity; in particular, computational creativity should include more than merely replicating norms, and completing independent tasks within the creative process (with little feedback or collaboration between tasks). The impact for computational creativity is that for these larger scale multi-system public-facing projects to be more successful, we are reminded of the need to develop as well as replicate human creative achievements, and to allow our systems to be able to communicate and refine work as well as offer inspirational material.

Introduction

In musical theatre, those involved in shows are accustomed to being reviewed. Essentially this paper provides a review of the 2016 musical *Beyond the Fence* - but a review with a difference. *Beyond the Fence* has been billed as “the world’s first computer generated musical”.² With several computational creativity software packages and computational data analyses providing data, frameworks and content for the musical in collaboration with (human) musical theatre experts,

¹<http://www.wingspanproductions.co.uk/news-and-awards/read/48/Beyond-the-Fence-the-world-s-first-computer-generated-musical> (Mar’16).

²<http://beyondthefencemusical.com/> (Mar’16).

Beyond the Fence tests the theory of whether computational creativity can be used to create a musical. This work was undertaken by the television production company Wingspan Productions (led by Dr Catherine Gale) for a two-part documentary about the process, commissioned by a UK satellite channel (Sky Arts) with support from Wellcome Trust.

At the time of writing, the musical is coming to the end of a two-weeks-long run in London’s West End, an area of London with an extremely vibrant theatre and musical theatre scene (to the extent that this part of London is colloquially referred to as ‘Theatreland’). Having been involved in *Beyond the Fence* as an informed commentator who was not involved in creating any of the software, I have had the opportunity to gather and discuss information about this project with a variety of different sources, from people behind the software to the cast performing the show. From this perspective, two interrelated questions have emerged:

1. To what extent is *Beyond the Fence* successful?
2. What does this contribute to computational creativity?

These questions guide this paper, in evaluating the *Beyond the Fence* project via each of these questions. Some details are given about how *Beyond the Fence* has been undertaken and what creative entities have been involved.

To evaluate the project in a computational creativity context and address the above two questions, it is treated as an example of interactive creativity and is evaluated using a framework advocated for this type of creativity (Kantosalu, Toivanen, and Toivonen 2015). During evaluation, in conjunction with personal communications with Gale (2016) on how to judge success, some questions are considered in this paper as possible metrics for gauging success. Accompanying this, various traditional metrics for success in musical theatre are explored below. Evaluation affords us insight on the extent to which this project has been a success, and the engagement of this project with computational creativity research.

The paper concludes with a discussion of what the field of computational creativity can learn from the *Beyond the Fence* project. Where has computational creativity successfully contributed to *Beyond the Fence*? What has not worked so well in terms of computational creativity’s application to this problem of creating a new musical? And where would future work in this direction be most usefully directed?

Details of the project

“September 1982. Mary and her daughter George are celebrating one year of living at the Greenham Common peace camp. The group of women they have joined are all committed to stopping the arrival of US cruise missiles through non-violent protest. When Mary is faced with losing her child to the authorities, an unlikely ally is found in US Airman Jim Meadow. How can she continue to do what is best for her daughter while staying true to her ideals? *Beyond the Fence* is a powerful new musical about hope, defiance, unity and love.”³

Another story of interest to a computational creativity audience, however, is not the end product itself but the process and interactions that took place between different creative entities to create this musical. This process was driven by a team from Wingspan Productions, who describe the engagement with computational creativity as follows:⁴

“The process began with a predictive, big data analysis of success in musical theatre, conducted by Dr James Robert Lloyd, Dr Alex Davies and Prof Sir David Spiegelhalter (Cambridge University). They interrogated everything from cast size, to backdrop, emotional structure to the importance of someone falling [in] love, dying (or both!) - in more and less successful shows - to create a set of constraints to which the musical had to conform, to theoretically optimise chances of success. Next, the team visited what’s known as the What-If Machine at Goldsmiths, University of London. With Prof Simon Colton, Dr Maria Teresa Llano and Dr Rose Hepworth at the helm, the machine generated multiple central premises, featuring key characters, for the new show. The team selected this as the starting point and the original idea for the musical: What if a wounded soldier had to learn how to understand a child in order to find true love?”

A plot structure for the musical was also generated computationally, thanks to work led by Dr Pablo Gervás (Complutense University of Madrid). A brand new analysis of musical theatre narratives enabled him to adapt an existing story telling computer system, called PropperWryter, to turn its hand to musicals and build the core narrative arc of the new show.

Taken together, all of the above enabled the precinct for the emerging story to be identified: Greenham Common. The team then wrote a book and lyrics (with the assistance of some other computational tools) that fitted all these constraints.

Finally, the music material has been provided by Dr Nick Collins (Durham University), who has created a computer composition system based on a machine listening analysis of musical theatre music, conducted by Dr Bob Sturm (QMUL) and Dr Tillman Weyde (City University). Additional computer music material [was] generated using the FlowComposer system created by Dr Pierre Roy and Dr Francois Pachet (SonyCSL, Paris).⁵

In the credits for the musical, the ‘creative team’ listing includes the software programs involved and key researchers

³<http://beyondthefencemusical.com/about-the-show> (Mar’ 16).

⁴A paper has been produced by the teams involved, giving fuller details of how *Beyond the Fence* was constructed. This paper is not intended to duplicate these descriptions, but to critique the project from an independent perspective.

⁵<http://beyondthefencemusical.com/the-science> (Mar’ 16).

on each piece of software, plus two human musical theatre experts (Benjamin Till and Nathan Taylor), who curated the software outputs into its final musical format. The accompanying documentary shows how human members of the creative team took care to adhere to the spirit of this project: using as much computer-generated material as possible even when this caused difficulties. As Neil Laidlaw (producer of the stage performance) says, “we have to honour what we’ve signed up for” (Wingspan Productions 2016).

During investigations, Gale and her team became interested in what ‘being creative’ actually means and how creativity might emerge from rules and be assessed, given the difficulties we have in assessing creativity in humans. One particular debate Gale looked at in her conversations with computational creativity researchers relates these thoughts directly back to *Beyond the Fence*: does a generative process have to result in a good quality product (i.e. the musical) to qualify that process as having been creative? This distinction between the process and the generated artefact when assessing/recognising creativity has often arisen in computational creativity research (Gervas 2009; Jordanous 2016, for example). For *Beyond the Fence*, then, can the project be considered successful even if the generated show is not well received as a musical in its own right?

Evaluation

To what extent is *Beyond the Fence* successful and what does this contribute to computational creativity? To evaluate both of these questions, we consider *Beyond the Fence* as a case of *interactive creativity* generating a musical. Evaluation models typically focus on evaluating a single system, but Kantosalo, Toivanen, and Toivonen posit the DECIDE framework (Rogers, Sharp, and Preece 2011) as a model of evaluation suitable for evaluation of several systems collaborating and being co-creative with humans; this is the scenario we have with the creative systems used for *Beyond the Fence*. Hence to evaluate the interactive creativity in this project, following Kantosalo, Toivanen, and Toivonen (2015), we use the DECIDE framework:

1. “Determine the goals
2. Explore the questions
3. Choose the evaluation methods
4. Identify the practical issues
5. Decide how to deal with the ethical issues
6. Evaluate, analyze, interpret, and present the data”

DECIDE: Determine the goals The Wingspan Productions team conducted this project to explore if a ‘computer-generated musical’ was possible. Specifically, their goal was to create and stage a musical generated in collaboration between creative software and human musical theatre experts.

DECIDE: Explore the questions As part of this project, the Wingspan Productions team sought out and engaged with several leading computational creativity researchers, as described above. During this process, the team grew more interested in various debates and issues around computational creativity, and how the *Beyond the Fence* project sits in the wider context of computational creativity. During the process, Gale and her team explored how the *Beyond the Fence*

related to various key areas of concern in computational creativity research. To gain better understanding of computational creativity research, specific questions emerged (Gale, 2016, personal communications):

1. “How has our attitude towards how we use computers changed in recent decades?”
2. Why do people develop machines that are creative?
3. So is it right to paint a portrait of machines as a young artist? One that is maturing?
4. What kind of systems have been/are being developed?
5. What different approaches do people take? ‘Heroic’ methods where computer is [an] artist in its own right, or more collaborative approaches?”⁶

In conversations with various computational creativity researchers, including myself, Gale also investigated questions about the role of computer software in the ‘creative conversation’: “Computers can become another voice in the room - speaking ‘from the data’ as it were - and we instinctively question that” (Gale 2016, personal communications). Gale observed through this experiment that people are often surprised at the challenges and difficulties in computationally generating creative artefacts - perhaps underestimating the complexity of the tasks involved. She also saw resistance in people’s reactions to computers being creative. One example of this resistance is illustrated in the documentary that reports on this project (Wingspan Productions 2016). Benjamin Till reflects a number of times on his apprehensions about working with computational software, such as this quote from his interactions with the products of the *Android Lloyd Webber* music-writing system. He says: “maybe I was a little bit harsher on it than I should have been”(Wingspan Productions 2016), going on to explain that because *Android Lloyd Webber* was computer software, he partly felt that he did not want the results to be good.

Concentrating on the *Beyond the Fence* musical project itself, Gale and her team were interested in what computational creativity researchers thought about the project; such discussions receive attention in the documentaries resulting from this project (Wingspan Productions 2016). For example, Gale was interested in whether the *Beyond the Fence* project was doing work that was in some way different to existing current work in computational creativity, or work that was exciting for the field. Relevant aspects that emerged in such discussions included the collaborative aspects of the creation of the musical, and the scale of the overall project (especially as the project resulted in public performances presented in a venue in a high profile London location.).

DECIDE: Choose the evaluation methods For this experiment, what will constitute a success? The question of whether there was an underlying hypothesis for the project was raised in a question-and-answer session post-performance I took part in with Catherine Gale (Wingspan Productions), Bob Sturm and Benjamin Till (respectively representing computational and human parts of the music generation team) on 2nd March 2016. This discussion revealed some difficulty in pinning down an exact scientific hypothesis for the project (one by which the project success

⁶Probably inspired by d’Inverno and McCormack (2015).

could be tested against and/or measured). However in this discussion and in other personal communications with Gale (2016), two different ways of considering the success of the project emerged: (1) traditional evaluation of success in musical theatre, via reviews and other metrics (given the information available at the time of writing); and (2) the wider contributions of the project to pursuits of knowledge, independently of the success of the musical.

The big data analysis by the Cambridge team distinguished between ‘hits’ (culturally and commercially successful), ‘flops’ (culturally and commercially unsuccessful), ‘critically acclaimed’ musicals (successful culturally but not commercially) and ‘crowd pleasers’ (successful commercially but not culturally). Commercial success is relatively easy to gauge after a show has been performed for a length of time: for example by looking at the amount of money a musical makes, the length of its run, the number tickets sold and any touring the show does. Typically, cultural value is more subjective and so trickier to measure objectively. While people involved in the project have reported that they felt the show was a success (Wingspan Productions 2016) (and Taylor, 2016, personal communications), empirical metrics are possible (Jordanous, Allington, and Dueck 2015). For musical theatre, cultural value can be measured via awards, press attention, reviews, influence on other shows, audience reaction, pick-up by amateur companies, location of show venue, funding, and other non-commercially measurable aspects.

DECIDE: Identify the practical issues Practical issues in the *Beyond the Fence* project poses an interesting challenge in terms of evaluation of the creativity of the software involved. Multiple software was used during this project, as well as interactions with creative people. Hence we could either evaluate the creativity of each constituent software-based part of the creative team, or focus on evaluating the overall collective project. As this paper’s aim is to evaluate the multiple parts of the project as a whole, it focuses on the latter aim; the task of evaluating individual software falls better to papers that report on individual software.

The Standardised Procedure for Evaluating Creative Systems (SPECS) (Jordanous 2012) asks researchers to (1) identify a characterisation of creativity by which to evaluate creative systems (2) derive standards or benchmarks by which to measure our systems, and (3) devise suitable tests to evaluate our systems against these benchmarks. We have a number of models available to us that we could use as our base characterisation of creativity for Step 1. For example, the components of creativity derived in (Jordanous 2012) represent a general basis for creativity. Alternatively, the FACE model (Colton, Charnley, and Pease 2011) analyses creative systems on their ability to use and/or generate new methods for using Frames (natural language descriptions of what the software has done), Aesthetics (measures), Concepts (underlying theory/ies that guide the creative process) and Example outputs. The Creative Tripod (Colton 2008) asks whether the system under question can be considered as a candidate for a creative system, through identifying the system’s ability to demonstrate skill, imagination and appre-

ciation. Ritchie has devised empirical criteria for evaluating creative systems (Ritchie 2007), though the criteria are less applicable for this evaluation as they are based upon ratings of the typicality and value of multiple artefacts produced by a creative system, whereas we only have one output to judge.

Considering the interactive nature of the *Beyond the Fence* project, it is difficult to generalise our currently available models across multiple systems acting collaboratively. There is also the question of whether to include human parts of the creative team. These issues make the above-mentioned existing evaluation methodologies and methods difficult to apply in this evaluation.⁷ One criterion of what makes a good methodology for evaluating creative systems (Jordanous 2014) is the usefulness of feedback. As we have seen, the *Beyond the Fence* project comprises multiple software, each tackling different types of creative task. This illustrates the vast and varied scope of creativity in musical theatre. For this current work, this evaluation is intended to uncover formative feedback for future development and recognition of this work's contribution to research.

DECIDE: Decide how to deal with the ethical issues Gale (2016, personal communications) has reflected on whether computers could (and should) be considered creative entities, at a level which is comparable or equal to humans. She discussed with various computational creativity researchers what might make people working in the creative industries more receptive or better suited to collaborate with computers than others, and questioned how to talk to creatives about their attitudes, perceptions and potential biases towards working with computers. Certainly this latter question deserves greater attention if computational creativity research is to reach a wider audience and broader range of collaborative partners. Some interesting points have already been raised which we can build upon such as how people may perceive the creativity of a system by looking for key aspects that a system should generate if it might be described as 'creative' (Colton 2008); and the role of people's reactions on interacting with creative systems, as a key contribution to that system's creativity (Gervas 2009; Gervás and León 2014; Jordanous 2016). Computational creativity researchers could also consider to what extent it is reasonable (or productive?) to attribute creative agency to a computer when featuring computational creativity software in public engagement activities, following discussions on creative agency and creative responsibilities (Maher 2012; Johnson 2014; Bown 2015).

Ethically, there are also challenges for people in experiencing *Beyond the Fence* as a computer-generated artefact - as we saw, many of the reviews mentioned a feeling of disconnect at times - something missing from the experience. This was discussed in the previous section. Another ethical issue relates not to the human participants in the creative team, but the computational participants. Is it fair to test computational systems at a professional level, where they

⁷Tackling the question of what is the most faithful model of creativity in this particular scenario would be interesting and challenging work to carry out as a follow-up project in its own right.

are being required to generate material at a standard which it takes humans years to reach (founded upon decades and more of human experience more generally?) Typically, evaluation of computational creativity systems has been undertaken on a more controlled and less professional level, away from the public eye - although there are notable exceptions to this as exemplified by the *Painting Fool* (Colton 2012) or the *Unnatural Selection* (Eigenfeldt 2015), both of which have recently 'participated' in public professional displays of their creativity (exhibitions and concerts respectively).

One other issue relating to the computational participants is the level at which we are evaluating them. Each software takes on a creative task assigned to it, but here we judge the overall success of the project rather than the success of each task. What we do not consider is the success to which individual tasks have been identified, and to what extent the software fulfils any original requirements. If poor decisions have been made and a vital part of the task overlooked, the computational participants may be judged more harshly as a result, even though they were not asked to perform this missing part of the task. The global focus of this current analysis can only touch on individual system issues; we leave more detailed analyses of success at individual tasks up to the researchers behind the systems involved. Here we focus on what we can learn from the project as a whole.

DECIDE: Evaluate, analyze, interpret, and present the data

As outlined above, we consider the cultural and commercial value of the *Beyond the Fence* musical, considering both the success of the show itself as a piece of musical theatre and by the contributions made to computational creativity.

Data on commercial benchmarks for success of *Beyond the Fence* is not yet available. However, from informal feedback, one of the creative team reported that the show was getting good audiences every night, which he had been able to observe since he had attended every show to date (due to having different guests coming to see every show - a mark of cultural interest in its own right). The creative team member was happy about this observation, particularly as the show was being performed in a London 'West End' venue.

In terms of cultural value, the timing of this paper also means it is not yet possible to reflect on awards, influence on other shows or whether the show is picked up by amateur companies. However we can already see that the show is being performed in (and supported by) the Arts Theatre, London, a reasonably well-known 'indie' theatre in a part of London strongly connected to theatre. The overall project was supported by the Sky Arts television channel as well as Wellcome Trust funding. As reported by *the Londonist* publication, Phil Edgar-Jones (Director of Sky Arts) was very positive about this "fascinating project that we're extremely proud to be a part of. At Sky Arts, we're always excited by innovation and this venture offers an intriguing glimpse into how technology is changing music evolution. Can an algorithm create music with all the humanity, emotion and drama that a person can bring? This question captivates us. We cannot wait to see the result."⁸

⁸<https://www.londontheatre1.com/news/121796/new->

An exhaustive search of Google results towards the end of the musical's run reveals reasonably extensive national/local press attention in the form of 20 reviews, from specialist musical theatre sources to national newspapers. The remainder of this section focuses on an analysis of key issues mentioned in these reviews. Figure 1 shows the most frequently-occurring words seen in this review corpus. A large proportion of these words relate to the content of the musical, as seen in reviews of more typical musicals. Reviews contain many comments relating to work by humans in this musical, such as the strong cast. However, words such as 'computer' and 'experiment' in this word cloud illustrate that these reviewers are well aware of the computational origins of *Beyond the Fence*; several interesting points are examined.⁹

Reviews of success As one reviewer comments, the *Beyond the Fence* project is indeed “[s]eemingly wanting to be judged as the output of an experiment rather than a ‘proper show’ ” (BroadwayBaby, see Table 1). While this seems like a criticism, it is actually not too far from the truth as an appraisal of the project's aims (Gale, 2016, personal communications). Typically, reviews focused on judging the show on its quality and fit as a “proper show”: “Computers can help write a musical, it seems, but they can't yet write a good one” (Engadget UK). “This show is as bland, inoffensive, and pleasant as a warm milky drink” (Guardian). The Londonist was a little more encouraging: “[w]hat's our measure for success? Well, the audience we saw *Beyond The Fence* with was applauding just about every number and was brought to tears by at least two of them.”

Several reviewers reflected at the experiment as a whole as part of their reviews. What's On Stage asked if computers can create more challenging material and concludes: “[o]ne day, maybe, but not yet. Not yet.” The Financial Times reviewer judges that “this bold experiment doesn't solve the many contradictions it throws up.” Similarly, the Reviews Hub says that “*Beyond the Fence* is an interesting experiment but it shows that computers are a long way off from creating a hit musical.” “Here, the puppet masters' digital strings are still a little too visible” (Musical Theatre Review). The West End Frame concludes that “as a theatrical event the show is remarkable. However, as a piece of theatre in its own right *Beyond The Fence* doesn't stand strong; it feels contrived and clunky.” But perhaps the computational creativity community can feel more heartened for future work here: “it is an interesting development in the intersection between theatre and technology that I suspect we haven't heard the last of” (‘There ought to be clowns’ blog).

The validity of co-creativity in *Beyond the Fence* “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

computer-generated-musical-beyond-the-fence-to-premiere-at-the-arts-theatre-in-2016/ (Mar' 16).

⁹For sources for each review quote here, see Table 1.

a thought-provoking novelty” (Telegraph). A number of reviewers commented on how human members of the creative team “are, essentially, curating and correcting the computers' output” (What's on Stage). Rarely if ever do the reviewers allow the software participants any creative agency or responsibilities for their output. Instead of being treated as co-creators in an interactive creative situation, computers are often portrayed in reviews as learners rather than creatives, whose work the human participants are being asked to bring up to professional level for the final product. In this project the computer software is not able to engage in revision, respond to feedback (particularly as musicals can change from night to night in response to feedback) - they provide material for people to curate. “[h]ere is where 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”¹⁰ (Engadget UK).

One notable and fascinating exception to this generalisation is by Musical Theatre Review: “‘What if a wounded soldier had to learn how to understand a child in order to find true love?’” was generated by WHIM, the “What If Machine”. And in tribute, Ako Mitchell's US soldier Jim rubs his thigh in pain occasionally - less wounded soldier, more bloke who should have done a few more warm-ups before exercising. There's an emotional need for him that is implied in WHIM's question that is not addressed here, leaving the show's central love story to feel a little anaemic.” In other words, software has supplied a creative idea to its human collaborators that is not used well, to the detriment of the overall effect of the show.

Too formulaic? “[A]ny show built according to a formula runs the risk of sounding, well, formulaic. *Beyond the Fence* doesn't avoid this risk: despite the cutting-edge technology involved in its creation, the show itself is middle-of-the-road and predictable.” (Financial Times). “Follow a formula and - who would have thunk it - you get something formulaic” (What's on Stage). Most of the reviewers criticised the musical for feeling too pattern-driven and predictable, rather than including content to challenge rather than replicate musical theatre. Several reviewers criticised the musical for its lack of avant-garde, challenging or ground-breaking content. “Picking through old shows for tricks evidently means the plotting is predictable and at times a bit shallow” (Londonist). The issue with this type of criticism is that the software involved was developed on the task of “replicating the past, not challenging it” (What's On Stage). “Nothing moved the needle. Nothing felt fresh.” (Engadget UK).

Where computer-generated aspects moved away from typical output, this also attracted criticism. For example several reviewers criticise the lyrics for being atypical - an ironic example of this is where during such criticism, a reviewer highlights the ‘We are Greenham’ song as one of the songs “that work” (Musical Theatre Review), even though that song consists entirely of computer-generated lyrics.¹¹

¹⁰A provocative request for future research?

¹¹It should be noted that this song's lyrical content was generated differently to many other songs, using corpus analysis of protest

Sometimes criticism of machine-generated content permeated into parts of the production humans had responsibility for: one reviewer suggested that the live band (of human performers, performing music that had been orchestrated by humans), “sounded extremely robotic and monotonous - it sounded as if backing tracks were being used” (West End Frame). The Financial Times said, though: “writers have been using formulae for years to make commercially-minded culture and so what difference does it make if it’s a formula developed by a machine? This is the main talking point of the show ... a debate that will run and run.”

Lack of context awareness? As the Guardian reviewer observed, “The software appears to have ... zero grasp of 1980s feminism and the Greenham Common women’s peace camp. A pity, because that’s where it’s set.” (Guardian). But is this a fair criticism, given that the Greenham Common setting and feminist themes were chosen by the humans in the creative team rather than via software? Certainly many of a younger generation of musical theatre professionals could also be criticised for not knowing about this particular event in UK history, and the software is given no chance to research these themes post-decision to use them for *Beyond the Fence*. Criticising the computer participants for not being more knowledgeable seems harsh. But should (and could) the computers have used more contextual awareness, to develop ideas based on contextual information available e.g. via web resources? It seems from a number of reviewers’ comments that this wider contextual knowledge was expected, for example where reviewers criticised the show for not engaging more with feminism issues, or debates about nuclear weapons that concurrently happened at the same time as the musical was being performed. More than one reviewer commented that they would have liked to see a plot centred around scenarios a computer might have some knowledge of (e.g. the Financial Times suggest “circuit boards in revolt”), though no follow-on comment considers how human audiences might perceive the results.

One area where the computational creativity software was roundly criticised was in the ability to understand content unfolding over time, in longer-term structures. “Even if they give you a stroke of genius, they can never follow that up... every thought is a new thought” says Benjamin Till, in (Wingspan Productions 2016), where Nick Collins also reflects that this is an area for further research.

Gimmick by boffins? Biases and preconceptions It was interesting to see several reviewers make negative comments to the effect that poor human-produced musicals appear as if they were written by a machine, e.g. “Plenty of musicals written by humans sound as if they have been composed by a machine” (Guardian). The Telegraph reviewer reported how “the evening somehow over-rode my default scepticism”; others made more negative comments about ‘gimmicks’ introduced by ‘boffins’. Interestingly, during development, the work-in-progress musical was performed to a test audience of regular theatre-goers who were unaware of the origins of songs from Greenham Common rather than the Cloud Lyricist.

much of the material being computer-generated. The audience reacted positively to the preview performance, but what is more telling is their reaction once they were informed about the computational work and its contribution to the musical (Wingspan Productions 2016). They reacted with stunned silence, followed by nervous laughter. Preconceptions about what computers can (and cannot) do are there to be dealt with, in computational creativity - the reviews here show that this issue should not be ignored when engaging with the public in computational creativity research.

Conclusions and future directions

“Beyond the Fence is conceived by computer and substantially crafted by computer.”¹²

The *Beyond the Fence* project has achieved the goal of staging a musical which has been generated through collaboration between creative computer software and human musical theatre experts. At the time of writing, the *Beyond the Fence* musical is reaching the end of its two-week run in London’s ‘Theatreland’ (the West End district of London, UK). The premise, plot elements, storyline, music and approximately a quarter of its lyrics were computer-generated, using various creative systems in conjunction with human experts. The project has been recorded in documentary form (Wingspan Productions 2016). The human creatives involved have reported that they feel the project has been a success, and the show has been performed to good-sized and receptive audiences each evening of its run. While “this bold experiment doesn’t solve the many contradictions it throws up” (Financial Times, see Table 1), it has made these “contradictions” and debates open and relevant for discussion among a much broader audience.

What can the field of computational creativity learn from the *Beyond the Fence* project? Where has computational creativity successfully contributed to this project? What has not worked so well in terms of computational creativity’s application to this problem of creating a new musical? And where could future work in this direction be directed?

In the short term, this project has played an important part in raising the profile of computational creativity research. This project has taken on an ambitious task, and has tasked computational creativity researchers with applying the software we create both as individual pieces of software and (importantly for computational creativity) in combination with other systems. While some work has been done in combining different creative systems for a broader perspective on creative tasks (Monteith et al. 2011, e.g.), the idea of different creative systems communicating and/or collaborating with each other (Corneli et al. 2015, e.g.) is an exciting area to look at (especially now many different creative systems have been developed and are potentially at our disposal).

The project led the Wingspan team to become interested in questions to do with creativity in different domains. Perhaps because of her biochemistry research background, Gale particularly focused on people’s perceptions of creativity outside of artistic domains, and cultural issues that may affect how we distinguish between creativity in scientific do-

¹²<http://beyondthefencemusical.com/about-the-show> (Mar’16).

Table 1: The twenty reviews of *Beyond the Fence* sourced via search, for analysis in this paper

Source	Link
NATIONAL NEWSPAPERS	
Guardian	http://www.theguardian.com/stage/2016/feb/28/beyond-the-fence-review-computer-created-musical-arts-theatre-london
Telegraph	http://www.telegraph.co.uk/theatre/what-to-see/beyond-the-fence-arts-theatre-review-computer-says-so-so/
Financial Times	http://www.ft.com/cms/s/0/5f993b32-dee2-11e5-b67f-a61732c1d025.html
Independent	http://www.independent.co.uk/arts-entertainment/theatre-dance/reviews/beyond-the-fence-arts-theatre-review-despite-my-reservations-i-was-won-over-a6900836.html
The Times	http://www.thetimes.co.uk/tto/arts/firstnightreviews/article4702133.ece
SPECIALIST THEATRE PUBLICATIONS	
The Stage	https://www.thestage.co.uk/reviews/2016/beyond-the-fence-review-at-arts-theatre-london-futuristic-composition-with-traditional-problems/
What's on Stage	http://www.whatsonstage.com/london-theatre/reviews/beyond-the-fence-arts-theatre_39847.html
The Reviews Hub	http://www.thereviewshub.com/beyond-the-fence-arts-theatre-london/
Musical Theatre Review	http://musicaltheatreview.com/beyond-the-fence-arts-theatre/
West End Frame	http://www.westendframe.com/2016/02/review-beyond-fence-at-arts-theatre.html
British Theatre Guide	http://www.britishtheatreinfo.info/reviews/beyond-the-fence-arts-theatre-12609
Theatreworld Internet Mag	http://www.theatreworldim2.com/#!/beyond-the-fence-arts-theatre/fcmry
West End Wilma	http://www.westendwilma.com/beyond-the-fence-review/
BroadwayBaby	http://www.broadwaybaby.com/shows/beyond-the-fence/710587
Carns Theatre Passion	http://carnstheatrepassion.com/2016/02/27/beyond-the-fence-arts-theatre-west-end-until-5th-march/
EVENTS LISTINGS WEBSITES	
Londonist	https://londonist.com/2016/02/computer-penned-musical-beyond-the-fence-reviewed
There Ought To Be Clowns	http://oughttobeclowns.blogspot.co.uk/2016/02/review-beyond-fence-arts.html
Time Out	http://www.timeout.com/london/theatre/beyond-the-fence
TECHNICAL BLOGS	
Engadget UK	http://www.engadget.com/2016/03/02/beyond-the-fence-computer-generated-musical/

mains compared to artistic domains (Gale 2016, personal communications). Although this current project concentrated on creativity in musical theatre, it will be intriguing to see the directions of any future projects by Wingspan Productions concerning computational creativity.

Returning to the current project under discussion, what contributions does this *Beyond the Fence* experiment make to the computational creativity field: currently and longer-term? And given directions in computational creativity research, what might this musical be like in a few years?

Discussions during filming centred around a key point in the generation of the musical: in Gale's words: "right now, [do] humans have to be part of a project like this? [Do we] need some people in the mix to curate? and to make choices? as currently the computers involved can't censor their output very well, and they don't talk to each other yet either!" (Gale 2016, personal communications). Recent computational creativity research has focused on how creative computer systems might be able to interact with each other, communicate and give each other feedback towards refining and developing their own creative work (Gervás and León 2014; Román and Pérez y Pérez 2014; Corneli and Jordanous 2015). One exciting potential area for future work is in using computational creativity to carry out this 'curation' step. Can computational creativity software curate parts of a musical (lyrics, plot, characters, emotional timelines) into a single production? To what extent is interaction with a human(s) necessary in this process? Responding to the criticisms raised in reviews that *Beyond the Fence* is not computer-generated, but 'computer-assisted' or 'computer-

initiated' (as discussed above): to what extent can computer software actually generate the full content of a musical? Could software do everything? And if several systems are involved, how can we best evaluate the results? Could social media comments also be harnessed for evaluation - perhaps as a way of garnering instant feedback which can inform software in making edits to the show for the next evening's performance? Or does this lead us into a trap where we judge the programs doing tasks set by humans, via opinions of the end result rather than the success at each smaller task (without evaluating decisions taken on how to break complex creative tasks into subtasks)?

In this current project, the computer software are for the most part passive participants, in a process curated by humans. Essentially, as discussed above, the computational participants typically contribute material that is shaped by the human participants in the creative team, and the human participants have the final say in what makes it to the 'final cut'. Perhaps, to paraphrase the title of this musical, the 'fence' in musical theatre represents the recognition of computers as genuine creative participants contributing to the creative process. In this interpretation, the *Beyond the Fence* project does not fully see creative software moving 'beyond' this 'fence'. But certainly the debate on computers being creative has been opened up to wider public scrutiny, a debate to which the project makes a significant contribution.

To summarise: the *Beyond the Fence* project has been successful at showing how existing creative software can indeed be used within the scenario of creating a and acceptable musical, which has been moderately well-received by most crit-

