A Speculative Exploration of the Role of Dialogue in Human-Computer Co-creation

Oliver Bown
Interactive Media Lab
University of New South Wales
o.bown@unsw.edu.au

Kazjon Grace
Design Lab
University of Sydney
kazjon.grace@sydney.edu.au

Liam Bray
Interactive Media Lab
University of New South Wales
liam.bray@student.unsw.edu.au

Dan Ventura
Computer Science Department
Brigham Young University
ventura@cs.byu.edu

Abstract

In this paper we consider the notion of *dialogic creative artificial intelligence* (DCAI) systems, where co-creativity between a human user and a computational system is supported through dialogic interaction. By dialogue we mean both traditional language-based communication for explicit critique and persuasion (dialogue *about* creative artefacts) as well as a broader potentially non-linguistic notion of dialogue that emerges through the exchange of suggestions and changes (dialogue *through* creative artefacts). To capture this, we define DCAI as occurring when both system and user are able to influence each others' creative objectives. The paper motivates our pursuit of dialogic interaction and provides some explanation of why we have defined it through its impacts rather than its mechanism of action. We provide two analyses to support our argument: an exploration of a commercial creativity support tool that has an extensive vocabulary for describing artefacts abstractly but does *not* meet our definition of dialogic interaction, and a case study of a creative interaction between two human professionals that exhibits dialogic interaction throughout. For the latter we consider how studies of human-human co-creation can offer non-obvious design concepts that might be applied to co-creative DCAI systems.

Introduction

The use of computational creativity algorithms to create applied usable co-creative systems is now a significant focus of research in the field, requiring a combination of computer science, design and creative practice-focused research methods. Dialogue, both overt linguistic discussion and non-linguistic exchange of concepts and artefacts, is a critical component of human creative collaboration. In this paper we explore how different ideas of 'dialogic interaction' might be realised in co-creative systems, both in addition to and instead of a more rigid graphical mode of human-computer interaction (Shneiderman 1983). The main purpose of this paper is to develop an understanding of the nature of dialogue as an interaction design concept, what design requirements it presents and what challenges it poses for the current state of the art.

We do this by considering, as small case studies, one existing piece of commercial music generation software, and one observational study of a commercial music production team working together in a collaborative creative process (without the use of AI tools). These small case studies are used to examine the gap between existing modes of human-computer interaction involving creative AI systems, and potential scenarios that may more closely resemble human-to-human co-creative activities. As such, this is viewed as speculative design research (Auger 2013), albeit without a design artefact. We attempt to abstract the features of both types of interaction such that we might consider a range of design possibilities (i.e., not necessarily fixating on imitating the human-to-human scenario).

We begin by outlining some expectations of what dialogic interaction might entail in creative contexts, laying the foundation for the development of co-creative agents exhibiting what we call *dialogic creative artificial intelligence* (DCAI). We then go on to evaluate an existing piece of generative music technology according to these expectations, considering how it falls short, but also how it comes close in some ways, to satisfying the requirements for DCAI. Following that, we present our observations of a real human-to-human co-creative activity that clearly exhibits dialogic interaction. Lastly, we piece these elements together to consider some within-reach possibilities for supporting DCAI.

A Vision for Dialogic Creative AI

Co-creativity, in the context of computational creativity research, describes a scenario in which a human and a computer collaboratively make contributions to a creative outcome. It is common to frame a vision of co-creativity with reference to human-to-human forms of collaborative creative production, where the computer becomes promoted from a "mere" tool to a creative partner (McCormack 2008). For example, Lubart discusses four human-to-human metaphors that help categorise how computers might become creative partners: the computer as nanny, the computer as pen-pal, the computer as coach, and the computer as colleague (Lubart 2005), each describing a different way that the computer makes contributions to the output. In some cases, the computer's output could clearly be called creative in its own right with its own original contributions, whereas in others it is focused on giving stimulating feedback to the human user, to drive their creativity further. This may involve critique (i.e. communicated justified evaluation), which is by some accounts at least as challenging as, and important to, the creative process as generation.

Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

The rise of commercially successful conversational interfaces has drawn attention to the potential for more fluid forms of interaction with creative AI systems that may support creativity in new ways, including co-creative interaction that better resembles human-to-human co-creation. In comparison to the potential of conversational interaction, our current predominant model of human-computer interaction for computational creativity is one in which we trigger a 'generate' action on a traditional GUI interface and some output is generated. We may then iterate the generation process, tweaking parameters and selecting presets or data sets to refine our search, but essentially still engaging in a graphical instruction-based interaction modality, where a GUI provides the means to operate a given system. It is easy to see why this makes it hard to consider the system as a particularly active agent in the co-creation process.

Relative to this, a 'new' conversationally-grounded model of interaction would step away from the limits of the GUI interface and employ the open-ended nature of conversational interaction to support something more closely resembling human-to-human co-creativity. This vision needs to be unpacked, as current state-of-the-art conversational interfaces are often very constrained and do not guarantee to offer greater open-endedness of interaction. It is not their conversational interface per se that would afford greater creative potential, but the capability of the underlying computational creativity technology, for which any number of interface designs might be suitable to exploit this capability.

We introduce the term **'dialogic'** to help isolate what qualities we are interested in, then, in terms of new forms of productive creative interaction. Using this term, we hope to avoid a simplistic interpretation of the term 'conversational', as in the use of spoken word or written text as an interaction modality. We note, however, that our use of the term is quite closely aligned to Gordon Pask's notion of 'conversation' theory (Pask 1975), with which we do associate some of the qualities of dialogic interaction—this is a potential source of confusion.

The core features of what we propose as dialogic creative artificial intelligence (DCAI) are that:

1. In an iterative process, the system has the potential to actively and positively influence aspects of the creative objective, and that
2. It is in turn able to adapt intelligently to changing objectives.

The first feature establishes that the human-computer system is a *loosely coupled* network of creative production, through the mutual influence of its components. We acknowledge that this process of co-influence technically applies to a vast number of existing examples of creative tool use: many researchers and practitioners have observed how the tools they use may dictate the nature of creative outcomes. This is captured in philosopher Lambros Malafouris' description of the coupled interaction between potter, clay and wheel: there is no predominant starting point in the causal interaction that takes place as the pot is shaped, hence no objective sense in which the potter is an isolated creative agent, distinct from the objects he interacts with. More specifically, in our definition we require the influence to be active and positive, meaning that the system's influence on the user takes place *by design*, not as a byproduct of some other aspect of its design.

Thus if a creative process results in settling on a particular output (say a generated melody), but only because the creator failed to get the system to do what she wanted, i.e., through compromise, then we should not admit this as either an active or a positive act of influence by the system. On the other hand, if the system stimulated us to think differently and accept a certain type of output through persuasion or novel insight, then we would consider this active and positive. We acknowledge that there are likely to be fuzzy boundaries to this distinction.

The second feature requires the system to make contributions that are context-aware and responsive to the iterative input from the user. We may say, vaguely, that the system is truly *engaged* (and engaging) in the dialogue. Here and in several of the terms above, we are not seeking exact definitions, nor aiming for 'operationalisable' definitions. All of the above must be evaluated subjectively.

As an aside, we speculate that the second property (adaptability) may be critical in establishing the first (co-influence): a system can only actively and positively influence a creative process through proper engagement with the human's changing objectives. Again, the notion of 'loose coupling' can be useful in thinking of how two actors combine into a single temporary creative agency in the formation of co-creative outcomes. All of this serves to say that our vision for future co-creative systems is one where they have greater co-influence on the creative process, and that in doing so the interactions we have with them can be considered more dialogic than tool-like.

Motivations for Designing for Dialogic Creative AI

Our focus on dialogic systems is motivated in part by the emerging interest in understandable AI, and in part by our understanding of creative processes. Understandable AI is important because people are known to be more inclined to engage with faulty or less-effective AI systems if they receive direct, understandable responses from them, when compared to highly optimised systems that are neither directly responsive nor understandable (Oudah et al. 2018). A lack of comprehension is known to lead to mistrust and disuse of systems (Hancock et al. 2011), which in a creative context we think means little to no adoption of our systems beyond our own tight community.

Elaborating further on what it might mean to productively influence a creative process, we note that creativity research has long examined the need for divergent, heuristic, search-based cognitive work. This is often by definition: many consider creative tasks as those for which a clear path to the outcome is not known, and some form of blind (or at least ambiguity-tolerant) search must be employed (e.g., (Simonton 2011; Perkins 1996; Sternberg and Lubart 1991; Amabile 1996)). In human-human co-creativity, the technique of 'brainstorming' epitomises this type of divergent search in the form of a formalised activity. Supporting 'expansive' search is one way in which a co-creativity tool

might be expected to positively influence an outcome, and we already have examples of such work in the literature. For example, Karimi et al. demonstrate the use of a co-creative sketching tool that aims to actively stimulate a greater diversity of outcomes, and hence greater 'creativity', achieved through the stimulation of novelty, through its own sketch suggestions (2019).

However, most of the above-cited models of the creative process also stress that divergent search is, in itself, insufficient; there must be a process of review and filtering that tests ideas against goals and may drive the rethinking of those goals. Various models of creativity also specifically address the need for convergence, either in terms of convergent forces that constrain and direct divergent search, or as a separate phase. Simonton discusses the sketches used by Picasso, leading to his masterpiece *Guernica*, which demonstrate the presence of divergent blind search (2007). If nothing else, convergence is implied in the fact that creative tasks result in specific outcomes, but some tasks might involve gradual refinement. A dialogic co-creative process could simply stimulate divergence and go no further, but we might define it as more complete if it follows through to the process of narrowing in on a final outcome.

This also suggests that our two properties are related through the idea of the co-creative partners being loosely coupled; convergence can be achieved through the dual actions of the computer influencing outcomes but also adapting in a skilful, context-aware way, the success of which would be in part due to its ability to support convergence where relevant.

For convergent, focused creative development in particular, but also for expansive divergent development to a lesser extent, we believe that the system must be good at adapting to context and be capable of explaining itself (or giving context to its actions).

We note that as well as helping discover a solution that fits preconceived criteria, expansive search can also involve arbitrariness, which helps frame future creative goals and set helpful, narrowing constraints. It can be good to 'just get something on the table'. For example, Stokes' research into creative practice shows how artists set themselves up with an individual style through the establishment of constraints, which can in theory be arbitrary, as long as they serve to isolate the artist's individual style (2009).

Situating Dialogue in Different Interaction Paradigms

As we note above, we do not think DCAI is necessarily bound to conversational interaction (i.e., natural language), but we do recognise that there is a natural fit here. We believe dialogue could be achieved in a range of interaction contexts as a way to exchange information *about* creative artefacts without necessarily exchanging information *through* them. One natural situation for a dialogic interaction would be a request-based scenario (Bown and Brown 2018), where the user requests the system to produce outputs ("give me a funky bassline") and then iterates in response to what is offered ("make it more syncopated", "more like that last bit"). This requires some shared understanding of the subject, which requires the system both to adapt its understanding and to communicate it. A good explainable system might provide context or even be persuasive: "what about something in the style of Stanley Clarke?", or "a more syncopated rhythm would make sense against this drum beat".

We also note that a more 'ambient' (Bown and Brown 2018) mode of interaction might also provide a solid foundation for dialogic co-creativity. The bassline could simply be playing along in its own track while the producer works on something else, adapting to other elements in the music, without much direct communication, but with contextual factors influencing how the bassline develops.

These different modes, and the different levels of initiative, encapsulation and agency they represent are important to consider in terms of the overall user-experience and acceptance of the system. Clippy, the notorious Microsoft Paperclip, was disliked by many users for its intrusiveness and misplaced confidence, but many of the services it offered are actually automated as standard in today's word processors (Maedche et al. 2016). Under the right circumstances the user's sense of the system's involvement and engagement might enhance the experience rather than be irritating or counter-productive. Such design considerations can help frame how forms of dialogic interaction can be useful in co-creativity.

Analysis of an AI Music Production Tool

With this notion of DCAI in mind we turn to consider an existing commercial AI music production tool in terms of how these concepts play out in a provisionally co-creative scenario. We have not attempted to set this up as a formal user study because, to our knowledge, the tool is yet to be adopted in real production scenarios, and we feel that this is necessary for a formal user study to be relevant. Instead we take a form of heuristic walk through the properties of the system in a similar vein to types of heuristic evaluation used in interaction design research (Nielsen 1995). The aim is both to provide a worked example of how co-creative agents can be more dialogic in their interactions, as well as to motivate why that would be a good idea.

The tool is Splash Pro, the first offering of an end-user tool from Australian Music AI startup Popgun. Splash Pro is a generative creativity support tool aimed at amateur music producers, which runs as a standalone piece of software for Mac and Windows and also as a plugin for a number of music production environments. It works in the symbolic music domain, for which MIDI is the standard. We examined the tool running as a plugin for Ableton Live. This analysis refers mainly to the design of the interface, not to the generative capability of the system, although we recognise that the two are intertwined.

Splash Pro has three generation modes—create, accompany, and blend—with three different choices of instrument: keys, bass and drums (vocals is also an instrument option, but only in 'create' mode). Create mode produces a single melodic line or chord sequence, bassline or drum beat, with options to choose a style, which varies according to the instrumentation chosen. Accompany mode produces an accompanying chord sequence, bassline or drum beat, given

a source sequence to accompany. Blend mode generates an original sequence, but does so as a musically valid interpolation of two existing sequences, selected by the user.

In all cases, the interface requires the choice of a musical key and a duration over which to generate (it may also make use of the music's time-signature, which is not an option in the Splash interface, but can instead be accessed by plugins from the host program, where it is a user-settable parameter. Ditto with tempo). It also offers the user the choice to input a four-chord progression over the given number of bars.

Further, for each combination of generation mode and instrument, the system offers a number of preset stylistic options, as well as a number of numeric parameters. For create mode, with bass, preset options include either genre-tags: electronic, hip-hop, pop, RnB and rock, with each genre having subcategories such as hip-hop:boombap and hip-hop:modern. Alternatively you can choose a "technique" instead of a genre: long notes, short notes, on-the-beat short notes, accent on beats 1 & 3, busy moving notes, plays towards end of bar, short notes with groove, assortment of grooves, funky groove, less funky groove, unstable funk, etc. In addition, a set of numeric parameters can be modified: variation, note length, root octave, velocity, swing, and density. For other instruments, the options are varied to reflect that instrument. For example, for the numeric parameters for keys, chordiness, timing and pitch range are added, and swing and density are removed. This gives Splash Pro and its users a strong and shared vocabulary for reasoning about musical artefacts, which we see as a starting point for dialogic interaction.

Splash Pro has a graphical user interface based largely on buttons, popup menus and sliders, meaning that it has a defined set of GUI operations and options, rather than allowing any kind of open-ended construction of structures or behaviours by the user—an "operation-based" interaction paradigm (Bown and Brown 2018), as opposed to a request-based or ambient paradigm. As such, it presents as a regular software tool rather than a creative assistant.

It would be easy to conclude that such a tool does not satisfy our definition of dialogic interaction. However, as a generative software tool, we note that the system is implicitly involved in autonomously making 'suggestions' (even though its outputs may not be conceived of as such). It is widely documented how such output, even if not gathered in a dialogic manner, can strongly influence creative outputs when used by creative practitioners. Thus, in order to consider a minimal form of dialogic interaction, it is interesting to consider what would be needed to have the system adaptively respond, and complete a loop of loose coupling.

One option for extending Splash Pro would be enabling the user to give feedback to the system, such as ratings that might feed back into a generative process as an error signal (for example with reinforcement learning or evolutionary computing). This would permit adaptivity in both directions: the system to actively influence the user, and the user to actively influence the system. Alternatively, it could be possible for the system to 'ambiently' adapt and respond (which may not be evident through the GUI): even if the user doesn't give feedback, the system may be able to glean information through their actions.

Without some form of these adaptive behaviours Splash Pro, despite its extensive descriptive vocabulary, would seem to fall short of our criteria for being a dialogic system. This is no slight on the system: we are not convinced any successful instances of this exist in current creative software. Yet we might go even further in admitting Splash Pro as a dialogic system, since it provides a means for a user to feed it existing musical input, and through this input, it actually has a much more complicated interface than what is visible in the GUI. To fully examine this interface requires considering how someone might feed input into it.

Furthermore, through iterative processes, the user might fill in the gaps in what might make up a true dialogic interaction. For example, if the user wants to explore the interaction between different styles, they can combinatorially explore this space by feeding outputs back into the system. They could use the blend tool to blend two outputs that come from the create tool, and they could also use the accompany tool iteratively through different styles to create original combinations. Our informal experiments with Splash Pro, as well as with similar tools such as Google Magenta's Ableton Live plugin suite, have indicated that iterations of the generation process, feeding back content into the system's input, could constitute what might be called a pseudo-dialogic process. We assume the system is not adapting its internal state, but from a philosophical standpoint, when the system is coupled with the user and the musical content being worked on, the musical input can be thought of as an external state.

Although there is no option for feedback and no evidence of ambient adaptation in Splash Pro, we can still consider how it affords a sense of dialogic interaction through the user's own exploration of its response to different inputs and getting to know its behaviour. In the simplest sense, this is afforded through the sense of suggestion given in the stylistic parameters, as well as the sense of adaptation as the user themselves adapts and gets to know the behaviour of each of the generation settings. Since these are likely to be very complex models with rich behaviour, this sense of adaptation may be quite significant, even if no actual adaptation is happening.

Thus a rich enough system that was not adaptive could still present a sense of dialogic interaction by enabling the user to systematically iterate and progress ideas through their use of the system, with the shared creative workspace playing the role of a pseudo-adaptive intermediary: this may be the case if the system has very rich behaviour or a rich space of options, combined with some means to explore it. Exploration by interaction here can include submitting different types of inputs to the system experimentally. Even simple pop-up lists of options might be structured to create rich interactive pathways through a co-creative process. The system may have a large enough set of 'canned' responses to user inputs that navigating its interface still gives the experience of adaptation and suggestion. These would not meet our criteria for being dialogic, but they may increase a user's sense of the agency of the system.

A well-furnished vocabulary of terms shared between user and agent is, we argue, not sufficient to make a system

dialogic—for that it needs to be adaptive as well as actively and positively influencing the user. But a dialogic interface could be built on top of the current Splash Pro UI to allow the user to explore co-creation through the composition of operations; composition here referring to the creative and possibly adaptive iteration of these operations. Thus the user could say, "could it have a bit more of a Latin feel?" and the system could potentially present two or three different ways in which that adaptation to user input could be offered, with explanatory output to support that. The user could give positive or negative feedback on any such suggestions, driving the system to adapt how it blends these processes, and the system could offer more persuasive suggestions, drawing on previous success.

A Case Study of a Professional Music Production Process

We now turn to considering a real human-to-human co-creative scenario, studied through observation. Since dialogic creative AI systems—at least as we're envisaging them—don't exist, we set out to investigate a collaborative creative session between two musicians in order to identify how their interaction was dialogic. We are not suggesting by doing so that interaction with DCAI systems should imitate human-to-human interaction. Instead, we seek inspiration from human-to-human interaction to seed design ideas and requirements.

Our subjects, Uncanny Valley (UV), are a commercial music production studio based in Sydney. They are responsible for producing cues for major television dramas, reality TV, news and sport titles in Australia. They are also collaborative partners in our research into usable computational creativity systems and interested in adopting creative AI techniques in their workflow.

We observed two UV artists in a collaborative composition session: a producer, composer and keyboard player (henceforth A), and a co-composer and guitarist (henceforth B). Their brief was to create an alternative version of a track already previously composed by them, and used for TV coverage of a popular sports show. The alternative version was being used to promote a specific event within the upcoming season. The main melody of the original composition was to be kept and the track reworked given a specific reference track from the popular music canon ("The Man" by The Killers) which was provided by the client. B elaborates that the aim of such reference music is not to "sound alike" but to "vibe alike" to this reference, with a focus on tempo, groove and sound (a 'big' sound in this case). This is a recurring topic for the team who are often given reference tracks in this way: "There's nothing wrong with mentioning references, to talk about where you're headed, what you're feeling, but you know, we are always advising people to chase the same feeling not the same song." This gives a neat description of the creative task at hand: they seek to imitate the mode, key and tempo, and a similar instrumentation of the reference in a new arrangement of the provided melody.

During the observed session, which covered about half of the entire production process, a guitar part (recorded audio), chord progression (recorded MIDI), bassline (recorded MIDI) and drum part (constructed MIDI and samples) were laid down. Our focus is on the dialogic interaction between A and B in the creative development of the guitar part, considering as a design exercise how B, as a collaborating musician, might be substituted by an AI system. We also consider some of the other stages observed in order to give more context to frame this hypothetical scenario.

In terminology elaborated on in previous work (Bown and Brown 2018), we could describe A and B as interacting primarily according to a client producer model—with B as producer (of guitar content) to match A's brief. This is a nested relationship; the team of A and B are in turn producers responding to an external client's brief. It is also not a strict relationship, with B participating in direction as well, with equal creative authority on the direction. In this sense, this establishes the kind of complex interaction that we might wish to imitate in a dialogic creative AI experience.

Participant A describes their standard studio process as being heavily focused on a MIDI production paradigm using software synths and audio effects, being quick and efficient and easily manipulable. They identify the guitar as a key exception to this rule, being expressive, quick and robust to record, and supporting a "big" sound, that is key to many Western music styles, including the style the team identified as necessary to meet the provided brief.

This section proceeds by talking through the creative process, noting key observations along the way.

Observation

A required B to play the guitar to create a strummed rhythm track that backs the main melodic line. However, the creation of the guitar part was more of an ideation that influenced the overall development of the musical structure and was foundational to the establishment of other parts.

In the first instance this was workshopped. A expressed a basic brief which was discussed. They then prepared a rough drum track which was described as a temporary guide for playing to (not to be used in the final version), which expressed the groove; A emphasising how critical it was to establish the groove right from the start. Here "groove" refers to the basic rhythmic expression and emphasis.

A elaborates on the use of exploratory search in the early stages of a track: "Writing a track from scratch is sometimes a different process, but it will always start with a jam, either on piano or guitar. Before we did [another track] recently, it started with a played piano riff and then we wrote the melody over the piano riff. Then after the song was pretty much written as an entire thing, that's when we started producing it. Whereas this is more producing, writing at the same time. I like to produce and write at the same time; it saves time."

A refers to the initial jams as "expression sessions". B elaborates: "I'm self-confessed not a master of any instrument, just sort of spitballing at the wall, isn't it? Mud at the wall" B adds: "I'll leave the room almost deliberately often so that I can come back with observational power, otherwise sitting here listening to everything; I'll come back and go 'Oh, it's great, but that's the hook!' It'll be like 'Really?

I thought this was the hook!' like 'No, that's the hook!' Because I can hear it from downstairs and it's been going around and around and I can hear people whistling it downstairs, whatever." This highlights the iterative, negotiated re-framing of the creative task: B even deliberately removes themselves from the room in order to force the adoption of a different perspective. Here, dialogue that influences the creative objective is not only something that happens, but something that is being actively sought.

A also elaborates further on the need for live variation: "if you've got that little error there or little timing thing or that little note that you play, little ghost note that you play that you wouldn't do by yourself then you've got an extra layer of goodness there to add into the final mix. That's what gives music it's x-factor sometimes a little bit ... I've just learned that over the years of producing, if there's enough of that random stuff, jammy stuff going on at the beginning, then at the end you'll have something a bit better."

Reflecting this, the duo play together when recording parts (A performing the main melody on a MIDI keyboard), keeping eye contact, talking and using hand signals whilst playing. This is described as being done to get the groove, as well as being efficient. There's an interesting dichotomy in the way they talk about efficiency: in the early stages of an expression session, novelty, divergence and "jammy stuff" are clearly valued, but as the process continues, the focus shifts to speed and predictability, often after a specific meta-discussion of the need to progress. We suspect this is a common component of professional creativity: iteration and exploration are necessary, but deadlines are always tight.

A says: "At the early compositional stage of a track and finding all the bits and pieces that work together, if every part is aware of every other part then you keep... I think the reason I was jamming with B is because he can play the guitar and I can't, and I wanted to find what keyboard part was going to work over the guitar part, and then once we'd recorded that we sort of went, what bass part's going to work with this? And with the groove. And often it'll be happy accidents with jamming as well."

As well as using expansive, ambiguity-tolerant and diverse search processes to develop themes, the team also elaborate on how this process feeds their ideas about the overall structure of the track. It is apparent that determining the structure depends on specific content elements, and that this is strongly interdependent—the specifics of the elements in turn depend on where they sit.

A says: "Sometimes it's the case of thinking about what we're going to produce in the future while we're writing the riffs to understand what we need to get... That's cool." Thus as they jam and isolate guitar parts they like, they collaboratively develop a concept for the overall structure. This is described as "groove 1, groove 2, chorus" at one point, but later there is agreement that the second section is a variant of the first: "Groove 1, groove 1 on steroids, and then chorus".

As the order of parts is locked down, the specific details of the guitar riffs are also rapidly honed. A small handful of iterations of each take ensues, sometimes full repeats, clearly converging on specific phrases. The transition between exploration and exploitation is fast, and often explicitly encouraged by B, who was expressing a desire for the session to be as short as possible.

Having laid down the main part, the focus was turned to more specific details, such as "turnaround" moments, variations on the main themes emphasising the transition from one section to another, which involve discussion of the context of the transition (from where to where).

Multiple parts involve more extended discussion between A and B, taking different positions on what should be done. One such question was whether the parts should play more in unison or against each other. Through this back-and-forth, the two musicians attempt to actively and positively influence each other, and this dialogue is occurring both through music and discussion: it is both communication-through-creation and communication-about-creation.

Discussion

The human-human example clearly satisfies our definition of dialogic interaction, beyond the simple fact that the participants are in conversation. B inputs content that is informed and influences A, but also adapts to A. The team are very quick to converge on a stylistic focus, which combines their musical expertise with their experience of working together. The team moves beyond issues of genre and style rapidly to focus on finer specifics. Once within this stylistic focus, there continues to be a lot of rapid interaction which takes place through talking, gesturing and playing. It is fast and exploratory. We would describe their search phase as exhibiting a rapidly iterated curve of expansion followed by convergence. We note that during the expansive phase, many factors are up in the air: the song structure has still yet to be decided, and part of the convergence involves fixing the context for which the parts are being made. The collaborators navigate this ambiguity by establishing a shared but abstract dialogue about the desired qualities of the music: they don't need to debate extensively what a "big" sound is, they are both familiar with each others' stylistic vocabulary, and being too precise too early would harm the exploration.

We condense this to three observations regarding how a generative system might better support a dialogic co-creative process in a commercial music composition context, which may carry to other areas of creative practice.

Firstly, *interdependence*: the compositional process involved a great deal of interdependence ("all parts of the track must talk to each other"). The overall track structure depended on the components that arose and vice versa; it was beneficial to have parts played together to capture the same groove, which was established early on, and to carefully consider what elements might move in tight unison.

Secondly, *agreement of purpose*: the use of a stylistic reference was made in very specific ways. Through their experience and shared understanding of the process, the team was able to move very quickly into a very specific and targeted ideation stage. There was substantial ambiguity in their representation of this purpose, which was not just tolerated but encouraged by both parties.

And thirdly, *search processes*: the team made early use of fast-moving, expansive, exploratory (yet very stylistically-focused) search, laying out lots of options, guided by rapid

dialogue. They invited happy accidents and were attentive to quick re-conceptualisations of their goals and expectations, for example stepping out of the room. There was also very rapid but necessary convergence through refinement, once a number of basic decisions had been made. These convergences were often explicitly discussed, a kind of meta-dialogue about the process that shows another critical role for dialogic interaction.

We hypothesise that making our co-creative systems more dialogic will help create that sense of shared narrative that is known to characterise great human collaboration (Gottschall 2012). This is supported by our observational study—the musicians quickly developed a shared, abstract description for their piece that was more story than tight categorisation. People exchange information, feelings, and intentions using narratives (Polletta et al. 2011; Bruner 2009), and our proposed dialogic approach to co-creativity seeks to leverage this idea. Recent preliminary work in the domain of human-machine teaming validates this approach by showing that when AI systems are equipped with a rudimentary form of narrative intelligence, they are more likely both to establish and to maintain cooperative long-term relationships with people (Crandall et al. 2018; Oudah et al. 2018; Goodrich et al. 2018).

Narrative impacts people's view of robots (Rosenthal-von der Pütten, Straßmann, and Mara 2017), including perceptions of their usefulness and behavioral intentions. For example, Mara et al. observed that introducing robots as fictional characters increased people's perceptions of a robot's usefulness and behavioral intentions (2013). When thinking about what will make co-creative systems more likely to be adopted beyond their own creators, building trust and perceived utility through narrative seems promising. As narrative is a field of computational creativity in its own right, it may even be possible to explicitly represent and reason about this shared narrative that emerges from dialogic interaction, using ideas like co-operative narrative generation (Pérez y Pérez 2015). Computational narrative might also offer some ideas about how to evaluate dialogues (Kypridemou and Michael 2014; Rowe et al. 2009; Wang, Chen, and Li 2017), an essential part of dialogic AI that we've not touched on at all.

Dialogic interaction also has clear overlap with the notion of explainable creative systems. End-user explainable AI (as opposed to explainable techniques aimed at engineers and researchers) is in its infancy, but as it develops it will quickly become a critical component of co-creativity. Without the ability to explain suggestions, it will be impossible for a system to persuade, critique, or justify its behaviour, all of which would help build trust and acceptance (Zhu et al. 2018). Explainability has even been proposed as a possible "meta-aesthetic", or fundamental drive, for creative agents, based on the notion that ineffable ideas—those that cannot be expressed, or at least cannot be expressed concisely—are ineffective (Bodily and Ventura 2018).

The notion we are trying to capture with DCAI—the idea that co-creativity interaction should be about a shared dialogue through and about the work—is not a new one in AI. Enactive AI proposes a focus on autonomy and adaptability through interaction (Froese and Ziemke 2009), based on the notion of enactive cognition. Enactive cognition posits that reasoning arises through intentional interaction with the environment, and places a great emphasis on communication, both linguistic and non-linguistic (Cuffari, Di Paolo, and Jaegher 2014). This has been discussed in the field of creativity research as "distributed creativity" (Sawyer and Dezutter 2009), the idea, reminiscent of Csikszentmihalyi's systems view (2014), that collaborative creativity is "noninindividualistic", and "emerges from the improvised dialogues of the group". Enactive cognition provides support for our notion that dialogues can occur through interaction, and therefore need not necessarily imply chatbots or voice interaction, saying that co-operation can arise without "high-level mental processes" like direct communication (Fantasia, Jaegher, and Fasulo 2014).

Conclusion

We have proposed dialogic creative AI, a class of computationally co-creative system that we hypothesise is worthy of further study. DCAI has two core features: both human and artificial agents must have the potential to actively influence the creative objective, and must be able to adapt their behaviour as those objectives change. We have based this definition on the outcomes of establishing dialogue, rather than by defining dialogue itself. This focus exists because we want to capture a broader range of communication than traditional direct exchange of language—we want to capture both dialogue-through-artefacts as well as dialogue-about-artefacts. In short: It's not sufficient to slap an explanatory chatbot on top of your generative model, although that sounds like an excellent starting point. Our definition contends that there must be two-way influence and adaptability. We believe focusing on dialogue construed this way is essential for building creative systems that will be more broadly adopted in the future. Making that a reality will mean grappling with a host of related ideas: explainability, framing, critique, initiative, intent, meta-aesthetics, and more.

Acknowledgements

The informal study of Splash Pro was assisted by Chloe McFadden, a student in the Bachelor of Media Arts, UNSW, acting as a research intern. This work was supported by the Australian Research Council Discovery Project grant DP200101059.

References

Amabile, T. 1996. *Creativity in Context*. Westview Press.

Auger, J. 2013. Speculative design: crafting the speculation. *Digital Creativity* 24(1):11–35.

Bodily, P., and Ventura, D. 2018. Explainability: An aesthetic for aesthetics in computational creative systems. In *Proceedings of 9th the International Conference on Computational Creativity*, 153–160.

Bown, O., and Brown, A. R. 2018. Interaction design for metacreative systems. In *New Directions in Third Wave Human-Computer Interaction: Volume 1 - Technologies*. Springer.

Bruner, J. S. 2009. *Actual Minds, Possible Worlds*. Harvard University Press.

Crandall, J. W.; Oudah, M.; Tennom; Ishowo-Oloko, F.; Abdallah, S.; Bonnefon, J. F.; Cebrian, M.; Shariff, A.; Goodrich, M. A.; and Rahwan, I. 2018. Cooperating with machines. *Nature Communications* 9, Article No. 233.

Csikszentmihalyi, M. 2014. Society, culture, and person: A systems view of creativity. In *The Systems Model of Creativity*. Springer. 47–61.

Cuffari, E.; Di Paolo, E.; and Jaegher, H. 2014. From participatory sense-making to language: There and back again. *Phenomenology and the Cognitive Sciences* 14:1089–1125.

Fantasia, V.; Jaegher, H.; and Fasulo, A. 2014. We can work it out: An enactive look at cooperation. *Frontiers in Psychology* 5:874.

Froese, T., and Ziemke, T. 2009. Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence* 173(3–4):466–500.

Goodrich, M. A.; Crandall, J. W.; Oudah, M.; and Mathema, N. 2018. Using narrative to enable longitudinal human-robot interactions. In *Proceedings of the HRI2018 Workshop on Longitudinal Human-Robot Teaming*.

Gottschall, J. 2012. *The Storytelling Animal: How Stories Make Us Human*. Houghton Mifflin Harcourt.

Hancock, P. A.; Billings, D. R.; Schaefer, K. E.; Chen, J. Y. C.; De Visser, E. J.; and Parasuraman, R. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53(5):517–527.

Karimi, P.; Maher, M. L.; Davis, N.; and Grace, K. 2019. Deep learning in a computational model for conceptual shifts in a co-creative design system. In *Proceedings of the 10th International Conference on Computational Creativity*, 17–24.

Kypridemou, E., and Michael, L. 2014. Narrative similarity as common summary: Evaluation of behavioral and computational aspects. *Literary and Linguistic Computing* 29(4):532–560.

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(4):365–369.

Maedche, A.; Morana, S.; Schacht, S.; Werth, D.; and Krumeich, J. 2016. Advanced user assistance systems. *Business & Information Systems Engineering* 58(5):367–370.

Mara, M.; Appel, M.; Ogawa, H.; Lindinger, C.; Ogawa, E.; Ishiguro, H.; and Ogawa, K. 2013. Tell me your story, robot. Introducing an android as fiction character leads to higher perceived usefulness and adoption intention. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 193–194.

McCormack, J. 2008. Facing the future: Evolutionary possibilities for human-machine creativity. In Romero, J., and Machado, P., eds., *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*. Springer-Verlag: Heidelberg, Germany. 417–451.

Nielsen, J. 1995. 10 usability heuristics for user interface design. https://www.nngroup.com/articles/ten-usability-heuristics/. Nielsen Norman Group.

Oudah, M.; Rahwan, T.; Crandall, T.; and Crandall, J. W. 2018. How AI wins friends and influences people in repeated games with cheap talk. In *Proceedings of the 32nd National Conference on Artificial Intelligence*, 1519–1526.

Pask, G. 1975. *Conversation, Cognition and Learning*. New York: Elsevier.

Pérez y Pérez, R. 2015. A computer-based model for collaborative narrative generation. *Cognitive Systems Research* 36–37:30–48.

Perkins, D. N. 1996. Creativity: Beyond the Darwinian paradigm. In Boden, M., ed., *Dimensions of Creativity*. MIT Press. chapter 5, 119–142.

Polletta, F.; Chen, P. C. B.; Gardner, B. G.; and Motes, A. 2011. The sociology of storytelling. *Annual Review of Sociology* 37:109–130.

Rosenthal-von der Pütten, A.; Straßmann, C.; and Mara, M. 2017. A long time ago in a galaxy far, far away...The effects of narration and appearance on the perception of robots. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication*, 1169–1174.

Rowe, J. P.; McQuiggan, S. W.; Robison, J. L.; Marcey, D. R.; and Lester, J. C. 2009. STORYEVAL: An empirical evaluation framework for narrative generation. In *Proceedings of the 30th National Conference on Artificial Intelligence*, 103–110.

Sawyer, K., and Dezutter, S. 2009. Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts* 3:81–92.

Shneiderman, B. 1983. Direct manipulation: A step beyond programming languages. *Computer* 16(8):57–69.

Simonton, D. K. 2007. The creative process in Picasso's Guernica sketches: Monotonic improvements versus non-monotonic variants. *Creativity Research Journal* 19(4):329–344.

Simonton, D. K. 2011. Creativity and discovery as blind variation: Campbell's (1960) BVSR model after the half-century mark. *Review of General Psychology* 15(2):158.

Sternberg, R. J., and Lubart, T. I. 1991. An investment theory of creativity and its development. *Human Development* 34(1):1–31.

Stokes, P. D. 2009. Using constraints to create novelty: A case study. *Psychology of Aesthetics, Creativity, and the Arts* 3(3):174.

Wang, T.; Chen, P.; and Li, B. 2017. Predicting the quality of short narratives from social media. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3859–3865.

Zhu, J.; Liapis, A.; Risi, S.; Bidarra, R.; and Youngblood, G. M. 2018. Explainable AI for designers: A human-centered perspective on mixed-initiative co-creation. In *Proceedings of the IEEE Conference on Computational Intelligence and Games*, 1–8. IEEE.