Issues of Authenticity in Autonomously Creative Systems

Simon Colton[1, 2], Alison Pease[2] and **Rob Saunders[1]**
[1]The MetaMakers Institute, Falmouth University, UK
[2]Computational Creativity Group, Goldsmiths, University of London, UK
[3]School of Computing, University of Dundee, UK
www.metamakersinstitute.com ccg.doc.gold.ac.uk

Abstract

As the value of computer generated artefacts increases, we need to question how creative software will fit into human culture, both as creative collaborators and autonomously creative entities. We hypothesise that, in certain contexts where creative software has increased or total autonomy, lack of authenticity will be a major limiting factor for creative systems in the arts and possibly elsewhere. After explaining and motivating this hypothesis, we study notions of authenticity in the human context and use this to raise issues of computational authenticity. We propose a number of ways to address these issues, including influencing public perception of creative software, practical approaches to software recording and using its life experiences, and employing alternative methodologies for building creative systems. We believe that identifying, exploring and addressing issues of authenticity will help to maximise the beneficial contributions that autonomously creative software can bring to society.

Introduction and Motivation

As with many other areas of AI research and practice, there has recently been somewhat of a step change in the quality of artefacts generated through the employment of advanced techniques such as deep learning, and in some cases similar advances in the ease of use in deploying generative systems. As an example, looking at artistic/textural style transfer, the process previously involved writing a bespoke program to apply a certain visual texture to an image, or to generate art in the style of a particular artist or movement. Now style transfer merely involves supplying a style image and target image and waiting while a generative artificial neural network is trained and employed, which captures aspects of both style and target in a new image (Gatys, Ecker, and Bethge 2016), with remarkably good results. While pastiche generation is not normally seen as particularly creative, deep learning researchers are beginning to advance the creative autonomy of their generative systems, for instance through so-called Creative Adversarial Networks (Elgammal et al. 2017), and Computational Creativity researchers are employing such techniques in existing autonomous systems such as The Painting Fool (Colton 2012).

As the ubiquity of creative systems and the quality of their output increases, we have to consider how it will fit into human society. Like any advance in technology or change in societal values, there will be pros and cons to having truly autonomous, creative, software embedded in human culture. We start from the position that the advantages of having such software will far outweigh any difficulties that it would bring. The Computational Creativity community is somewhat split over the question of whether effort should be spent in advancing software to fully autonomous levels of creativity, or whether it would be better to concentrate on building co-creative tools for mixed initiative usage. We focus here on future scenarios where software can exhibit the same level of creative autonomy as people in the arts and sciences. In this context, we question whether opportunities for creative systems (and attendant advantages to human society) will be missed through a lack of authenticity in software, due to fundamental differences between people and machines, including a lack of life experiences to draw upon.

To motivate studying authenticity as an issue, consider the following demonstration, which has been carried out at more than a dozen talks by the first author, known hereafter as the *presenter*. The poem in figure 1 is presented as a well-known piece by female poet Maureen Q. Smith. There is then some joint discussion around what the author may have meant with this short poem about childbirth, for instance: the poem may be about her own experience, as per the 'My boy' sentence; the 'begin again' phrase may have been a reference to a literal beginning (of life), or a re-boot for the family; the 'joy', 'pain' and 'tears' are probably literal; the 'fears' may be about the birth process, but equally about the future of the world the baby is born into.

The presenter then points out that he has made a mistake: in fact, the author was a man, called Maurice Q. Smith. A re-evaluation is then undertaken, with the 'pain' perhaps now being projected, or expressing a painful worry about the

Childbirth

Maureen Q. Smith

The joy, the pain, the begin again. My boy.
Born of me, for me, through my tears, through my fears.

Figure 1: Poem about childbirth used in a demonstration addressing authorial intent and authenticity.

birth process. The presenter then points out that Maurice Q. Smith was actually a convicted child molester when he wrote the poem, and that it was widely regarded as depicting the process of grooming a child. The following re-evaluation highlights the suddenly much darker interpretation of 'Born of me, for me' with aspects such as 'The joy, the pain' becoming disturbing if taken literally, and 'fears' perhaps portraying worries about being caught in a criminal act.

The presenter then says that the poem was actually generated by a computer program using texts about childbirth, and suggest that – with no concerns of an unsettling backstory – another re-evaluation can take place. At this point, the demonstrator attempts to hypothesise what the software *meant* when it wrote about 'joy' and 'pain', but realises there was no further valuable meaning to be gleaned from an understanding of the generative process. Similarly, while the 'tears' are obviously not literal, it's seems impossible to project any further meaning onto the usage of this term. The presenter then invites the audience to question whether the poem has now lost some or all of its meaning, whether – now we can't indulge in projecting thoughts/experiences/intentions onto the author – the value of the poem has decreased or not. There tends to be agreement that people project more value onto the poem when they believed it was written by a person. For full disclosure, the presenter ends by revealing that he wrote the poem for the purposes of the demonstration.

Of course, poems can be – and often are – written, and read from third party perspectives. Well-intentioned ideals such as *Death of the Author* (Barthes 1967) and the *intentional fallacy* (Wimsatt and Beardsley 1954), advocate taking poetry and literature at face value, without inferring authorial intention and actively ignoring any knowledge of the author. It is not clear how easy such ideals are to implement in practice. The childbirth example shows how knowledge of merely the author's name naturally affects the reading of the poem, providing much context. Also, poetry is often enjoyed as a performance art, e.g., in poetry slams, where the personality and intentions of the poet shape the performance, and are explicitly included in commentaries. In a more reasonable appreciation of poetry, where at least some people infer authorial backgrounds and intentions – as highlighted by the childbirth poem – there may be an *uncanny valley* effect (Seyama and Nagayama 2007), e.g., if software autonomously wrote a beautiful ballad about teenage love, audiences would question what the software really knows about this topic, and hence what value the song really has.

We believe that many issues around lack of meaning and authorial intention can be understood via the lens of *authenticity*, or lack thereof, in computational systems. In the next section, we study authenticity from various perspectives in a human context. Following this, we use the study to raise certain issues about software authenticity, and suggest ways to address some of the issues, broadly in three areas: managing public perception of software authenticity; enabling software to use its life experiences in the creative process; and employing alternative methodologies for building creative systems. We conclude by discussing what these issues may mean for the future of Computational Creativity research.

Authenticity in the Human Context

A Million Little Pieces by James Frey (2003) is an autobiography about a struggle with drug addiction and rehabilitation. Published in twenty-nine languages, it has sold over 5 million copies, was on Oprah Winfrey's Book Club selection and number one on the New York Times Best Seller list. In 2006, The Smoking Gun published an article (thesmokinggun.com/documents/celebrity/million-little-lies) claiming that many events in (Frey 2003) had not happened and that Frey had fictionalized his life. The public took this hard: Oprah Winfrey said she felt "duped" and publically rebuked him for "fabricating" and "lying" and the public felt "betrayed" (Wyatt 2006). More than 10 class action lawsuits were filed on the grounds of negligent misrepresentation and consumer fraud, with readers asking for compensation for the time they "wasted" reading a book they thought was non-fiction. Publisher Random House withdrew from a deal with Frey and offered full refunds to readers, and some libraries re-catalogued Frey's book as fiction. Post-2006 editions come with disclaimers by both publisher and author, in which Frey writes:

> "My mistake, and it is one I deeply regret, is writing about the person I created in my mind to help me cope, and not the person who went through the experience. . . . I believe, and I understand others strongly disagree, that memoir allows the writer to work from memory instead of from a strict journalistic or historical standard. It is about impression and feeling, about individual recollection . . . It is a subjective truth, altered by the mind of a recovering drug addict and alcoholic. Ultimately, it's a story . . . that I could not have written without having lived the life I've lived." (Frey 2006, p2)

The debate and strong feelings about *A Million Little Pieces* centre on our modern notion of authenticity. This is an ethical characteristic, an ideal which shapes our worldview – "that one should be true to oneself and lead a life that is expressive of what the person takes herself to be" (Varga 2013, p.5). Authenticity is particularly valued in today's "post-truth" culture where the "fake" can spread widely and impactfully via social media and other channels; where we are urged by the self-help movement to get in touch with our "authentic selves"; and where perception of "brand authenticity" is thought to be the prevailing purchasing criterion of consumer behaviour (Morhart et al. 2015). Varga (2013, p. 5) proposes that we are living in "the age of authenticity", and Wilde (p. 361 of (Lindholm 2013)) writes that:

> " "Know thyself" was written over the portal of the antique world. Over the the portal of the new world "Be thyself" shall be written."

We consider here approaches and responses to authenticity in human creativity, from perspectives of Western philosophy, aesthetics, literature, empirical psychology, consumer behaviour research and cultural tourism. Philosopher of art Denis Dutton (2003) provides the useful distinction between *nominal authenticity* – establishing provenance of an artefact – and *expressive authenticity* – whether an artefact genuinely reflects an author's beliefs and values in a socio-historical context. Here, we focus on the latter.

Acceptable Inauthenticity and Non-authenticity

Socially acceptable levels of inauthenticity in human authors and artists vary, depending on culture, audience, time, and the authors themselves. One of the first English novels, "Robinson Crusoe" was presented as an autobiography of a sailor who was stranded on an uninhabited island in the Caribbean for twenty-eight years. Defoe included the phrase "Written by himself" on the cover page, and wrote – as editor – in the preface to the first edition: "The Editor believes the thing to be a just History of Fact; neither is there any Appearance of Fiction in it." In reality, Defoe was a journalist who never left Europe, although he knew of several real-life survival stories. While Defoe did face some criticism at the time, being called "a false, shuffling, prevaricating rascal" by Joseph Addison (Baker 2009), his pretense did not seem to affect the popularity of the book. This is perhaps because the genre of the novel was newly emerging and rules of convention had yet to be formed. Contemporary critics convey contemporary expectations, such as Nicholson Baker (2009), who calls Crusoe "Defoe's most famous hoax".

Since Defoe, there are many examples of people writing *as though it were written by someone else*, with varying degrees of deception. To our knowledge, there has been no criticism of contemporary author JK Rowling for being inauthentic, although she does present us with several examples of authenticity/inauthenticity in her writing. As an example, her wish to be known by initials was partly so that boys might assume she was male, as she thought they would then be more likely to read her stories (note that JK Rowling has no middle name: the 'K' comes from her grandmother's Christian name). Moreover, after success with the *Harry Potter* books, Rowling adopted the non de plume Robert Galbraith, and further invented an appropriate persona and a fictional biography directly related to Galbraith's story topics, adding depth and credibility. This was presented on "his" early books as fact, e.g., the author's biography on the inner sleeve of (Galbraith 2013) reads:

> "After several years with the Royal Military Police, Robert Galbraith was attached to the Special Investigative Branch ... He left the military in 2003 and has been working since then in the civilian security industry. The idea for Cormoran Strike grew directly out of his own experiences and those of his military friends who returned to the civilian world. 'Robert Galbraith' is a pseudonym." (Errington 2017).

Both of these aspects seem to be acceptably inauthentic, possibly due to Rowling's popularity, reputation, longevity and backstory, and to the fact that she made no further pretense that these were Robert Galbraith's experiences. This is in contrast to Frey, who was unknown before his book, and who promoted it with public appearances in which he propagated the persona he had created (albeit unconsciously). Of course, writing about wizards and witches might also have laid Rowling open to accusations of inauthenticity, but as she doesn't claim to have experience of them, there is no deception, hence this is better described as "non-authentic" behaviour, i.e., not on the authenticity/inauthenticity scale.

As another example, Mark Haddon, in his book *The Curious Incident of the Dog in the Night-time*, took a more transparent approach than Galbraith. The book was written from the perspective of a 15 year-old mathematician with behavioural difficulties. Haddon himself has none of these characteristics, but he escapes charges of inauthenticity by being completely open about who he is. Thus he also falls into the realm of the "non-authentic". These examples show different levels of fiction, ranging from the realistic fiction of Defoe's, Galbraith's or Haddon's books – stories that *could have* occurred, resembling real life, believable settings and characters, to the fantasy of the Harry Potter series – a form of speculative fiction, which also includes science fiction, superhero fiction, science fantasy, horror, alternate history and supernatural fiction. Such a scale may have implications for authenticity judgements, for instance with greater expectations of authenticity in realistic fiction than in speculative fiction, especially if a fictionalised world differed significantly from standard human experience.

The Problem with Experience, Memory and Self

Whatever the level of fiction, boundaries between fiction and reality are impossible to define. No fiction is entirely fictional, as materials are taken from reality. Conversely, reality is not all that solid, there is no formal, ideal perception and any representation involves a point of view, a perspective, criteria of relevance, an implicit theory of reality, and so on. Even if there were an ideal perception, creative representations are usually based on memory, which – as argued by Frey in his disclaimer above – is malleable and unreliable. Experimental psychologist Elizabeth Loftus and colleagues have shown that memories can not only shift, or be lost via decay and repression, but that new, false memories can be implanted, or "recovered" through therapy. She writes: "Only the flimsiest curtain separates reality from imagination"; "The detail people confabulate and then believe in just astounds me"; "What we access is halfdream, half-construct, entirely unreliable" and describes artefacts based on such memories as: "authentically inauthentic" (Slater 2004, Chap. 8).

Loftus shows us the fragility of the connection between our experience and our awareness of experience. This connects to a model of authenticity from person-centred psychology, which Wood et al. (2008) describe and use to develop an authenticity scale questionnaire, designed to show degrees of three different aspects of authenticity. These concern: (a) primary experience, involving physiological states, emotions and cognition; (b) awareness of experience, and (c) behaviour and emotional expression. The first aspect of the model of authenticity is called *Self-Alienation* and describes the relationship between (a) and (b), concerning mismatches between actual experience and conscious awareness of it. The second aspect concerns the relationship between (b) and (c), called *Authentic Living*, and involves congruence between experience as consciously perceived and expression of emotions. The third aspect, *Accepting External Influence*, involves the influence of social environments on both the first and the second aspects, as accepting influence of other people and conforming to expectations of others. Wood et

al. propose evaluating each authenticity aspect through a set of similarly phrased statements, with which participants express agreement or disagreement using a Likert scale, from 1 *does not describe me at all*, to 7 *describes me very well*. *Self-Alienation* is judged via statements like "I feel out of touch with the real me", *Authentic Living* via "I always stand by what I believe in", and *Accepting External Influence* via "Other people influence me greatly".

Varga (2013, page 61) elaborates our sense of self, assumed in the *Self-Alienation* aspect of Wood et al. (2008), describing two competing models. The *Inner Sense Model* assumes that the inner self is something which is stable and given, that can be known and expressed. The *Productionist Model* on the other hand, takes the position that there is no fundamental, unchangeable self comprising 'psychological DNA' to discover, but rather we continually create and re-create a dynamic, fluid self in different contexts. These ideas build on philosophical notions of personal identity, which explore issues such as defining and undefining features, changing and contingent defining features, personal identity over time, how we know who we are, how disunity of consciousness or split personalities affect our notion of synchronic identity and what is important when we think of a self. The question of what and who a self is, is important if by authenticity, we mean being true to ourselves.

The existentialist movement in philosophy has elaborated the notions of self and authenticity, seeing the experience of authentic feeling as validating our existence. Much of this was a response against excessive social influence, mannered falsity, refinement and hypocrisy – the *Accepting External Influence* aspect of the person-centred model of authenticity. Rousseau complained about people striving to impress each other and only being able to experience themselves as reflected in the eyes of others, being proud if admired, self-despising if held in contempt, etc. Nietzsche also reacted against his bourgeois, Christian upbringing, in which "no one dares to appear as he is, but masks himself as a cultivated man, as a scholar, as a poet, as a politician" [see page 383 of (Lindholm 2013)], developing the ideal of expressive authenticity and urging people to be themselves. Heidegger wrote about the inauthentic self, which "lives in subservient and unconscious relation to the anonymous and ubiquitous 'they"' [page 385 of (Lindholm 2013)]. Sartre further developed ideas of inauthenticity and bad faith, where identification with social roles, for instance that of a waiter, kills any possibility of authenticity. He writes: "I had been convinced that we were created for the purpose of laughing at the act we put on for each other" [page 387 of (Lindholm 2013)]. These ideas also have found traction in psychotherapy.

Expressive Authenticity in Political Sub-cultures

Expressive authenticity is an important concept in many domains. For instance, immersive musical sub-cultures such as the punk, goth and hip hop communities sometimes dismiss "part-time wannabes" or "hangers on" as poseurs, people assuming a persona in order to be accepted by or seen as members of a group, but who do not understand the group's values or philosophy. Here, we see a collective "self" in terms of authenticity, with members protecting their identity and valuing their authenticity, sometimes to the extent that people who do not share their collective experience are derided for producing – or even listening to – a particular style of music (Jacobson 2018). Personas of "poseur" musicians are sometimes created or elaborated by others, e.g., record companies and trade magazines promoting music stars. This highlights that a key factor in people's willingness or lack thereof in listening to an "inauthentic voice" is politically motivated. Someone from a privileged background, possibly capitalising on or stereotyping a politically marginalized group, may be far less palatable than the reverse. This accounts for the outrage at the James Frey example and possibly for any distaste or discomfort that neurotypical Mark Haddon wrote about a protagonist with autism.

Conversely to the importance of expressive authenticity given to the musical sub-cultures described above, some communities prefer to present artefacts as self-contained, stand-alone objects, independent from their creators. Presenting them alongside a view of the creator is thought to constrain the way in which they are interpreted and understood, putting artificial limits on the artefact. These ideas are advocated by New Criticism, a movement in literary theory, and Roland Barthes in his essay on "The Death of the Author" (Barthes 1967). Since we cannot understand authorial intent, and if we could, it would only limit our reading of their work, the argument is that literary fields should move closer to scientific or engineering fields in how they view creators. Once creators have brought an artefact or idea into being, their role is complete, and the artefact must stand alone and either work or not, much as we might judge a pudding or a machine (Wimsatt and Beardsley 1954, chap 1).

Brand Authenticity

Morhart et al. (2015) considered brand authenticity in the context of consumer behaviour, drawing on the literature and conducting exploratory interviews to develop four dimensions of perceived brand authenticity (PBA): continuity, credibility, integrity and symbolism. Here, continuity relates to longevity and persistent values; credibility to a brand's willingness and ability to deliver on its promises, which Morhart et al. conceptualise as transparency and honesty; integrity to virtue reflected in the brand's intentions and in the values it communicates; and symbolism to brands that reflect values that consumers consider important and that help to construct their image of themselves. Regarding authenticity, Morhart et al. summarise: "PBA is the extent to which consumers perceive a brand to be faithful toward itself (continuity), true to its consumers (credibility), motivated by caring and responsibility (integrity), and able to support consumers in being true to themselves (symbolism)."

The notion of brand authenticity is connected to cultural tourism, which is also heavily invested in the notion of authenticity. "Staged authenticity" (MacCannell 1973) consists of packaging or performing a cultural event in such a way that it conforms to expectations of an authentic tourist experience. The choice of venue and the surrounding context will all contribute to perceptions of authenticity as a consequence of the experience. Alternatively, cultivating a Benjaminian 'aura' (Benjamin 1968) through 'distance' can

enhance perception of authenticity in the arts. The emergence of authenticity as proposed by Cohen (1988) suggests that the inauthentic can become authentic over time, possibly as a consequence of the evolution of old traditions, or the establishment of new traditions. As Glaveanu (2017) points out, traditions are not fixed, but amorphous and shifting in response to cultural change. Finally, as Wang (1999) notes, the use of traditional practices creates a link to the past, relating the creative activity to the self, society and world.

Summary of Discussion

Up to this point, writings on authenticity have solely focused on its meaning in a human context. We have seen that:

- Evaluations of authenticity and its importance change over time, given different actors. Today, we may be more generous or forgiving in our authenticity evaluations of authors and artists who have built up popularity, reputation, longevity and backstory.
- Inauthenticity can take the form of obfuscation of identity (cf JK Rowling introducing ambiguity over her gender).
- Writing as though you were someone else can be acceptable. Acceptability is greatly enhanced if the author is transparent about it, where levels of transparency vary from producing a detailed fictional backstory and that role (possibly considered to be unacceptably inauthentic), to a non de plume and un-detailed persona but no accompanying appearances (possibly considered to be acceptably inauthentic), to complete transparency (non-authentic and acceptable).
- In fiction, it *may* be easier to avoid charges of inauthenticity if writing in "non-authentic domains", such as speculative, rather than realistic fiction (although there are contexts in which this wouldn't be the case).
- The notion of "being true to oneself" is psychologically and philosophy very difficult.
- There is a strong connection between life experience and behaviour or artefacts in expressive authenticity. In person-centred psychology, authenticity is described as the connections between *(a)* experience, *(b)* awareness of experience, and *(c)* behaviour which reflects that awareness, and the degree to which external influence affects these connections.
- The self can be seen as stable and knowable (discovered), or as dynamic and context-dependent (invented).
- In highly politicised domains, such as some sub-cultures of music, expressive authenticity is highly prized. In other domains, such as science, maths, and some literary and artistic approaches, authenticity is seen as less important.
- Perceived brand authenticity can be seen in terms of continuity, credibility, integrity and symbolism.
- Perceived authenticity of cultural events and artefacts can be affected by the way in which they are presented.
- Maintaining distance between artist and audience can build auras which might enhance perception of authenticity.
- The perceived authenticity of an artefact can change over time, as the cultural tradition and socio-historical context in which it exists also changes and evolves.

Dealing with Issues of Authenticity

The study above gives us a firm foundation on which to highlight issues of authenticity in software, and to suggest potential ways for dealing with valid related criticisms concerning autonomous software creativity. Anthropomorphisation has been both beneficial and detrimental to Computational Creativity researchers in presenting their work: on the one hand, it certainly helps to describe creative software in terms of human creative processes and products, but on the other hand, it can give a false impression of humanity when there is none. We have argued in (Pease and Colton 2012) that Turing-style tests can be detrimental to the bigger picture of embedding creative software in society. Moreover, in (Colton et al. 2014), we point out that – rather than levelling the playing field as hoped – such tests can actually serve to emphasise a *humanity gap*, i.e., people like an anonymised artefact because they make a human connection, but this is disappointingly removed on revealing that an artefact was made by computer, leading people to realise the implicit expectation of human creativity in making certain forms of art, such as poetry.

Addressing why such a humanity gap may be disappointing, we hypothesise that it is a perceived lack of authenticity as per some of the contexts given above. This can raise a dilemma amongst people appreciating computer-generated art: they may want to express dislike of a piece because of lack of human authenticity, but that may offend their liberal sensitivities in the context of the *Death of the Author* ideology described previously. That is, they may feel that they are pre-judging software unfairly in a way similar to racism or sexism, and it may not be clear to them whether this is a sensitive issue when applied to machine creativity. As more people are exposed to more high-quality computer generated artefacts, we believe this issue will become more pressing.

We advocate managing people's expectations of forming human connections when presenting computer generated material for cultural consumption, similar to someone being clear they have purchased an e-book rather than printed book for someone (Colton et al. 2014). That is, by escshewing Turing-style tests, getting software to frame its work, and being clear about the computational origins of generated artefacts, it seems possible to present computational creativity as being non-authentic rather than inauthentic, borrowing the terminology from the previous section. This could be taken further, i.e., by enabling software to **own its non-authenticity**, by it being clear that it doesn't have the relevant life experiences to bestow authenticity onto its process and product. The software could then suggest that audience members read/view/listen to its output as if it were created by a particular type of person, e.g., a teenage boy, or a particular (human) individual, etc.

Owning non-authenticity is a short-term possibility for side-stepping issues of authenticity. Another possibility is to **emphasise the product**, e.g., get the software to work in domains where product is far more important than process, e.g., scientific discovery. Put bluntly, if software invents a new cancer drug, no-one will care that it hasn't lost a relative to the disease. In the arts, abstract art is often perceived more as an invitation for a viewer to self-reflect than to inter-

rogate authorial intention, and here, authenticity may be less important. Moreover, as discussed above, authenticity may be less of an issue if, rather than working on realistic fiction, which may need human authenticity to support it, software instead **produces speculative works**, such as science fiction, which was the approach with The WhatIf Machine project (Llano et al. 2016). We note that even here, there may be contexts where authenticity would be paramount.

Referring to brand authenticity, we note that the authors of systems such as Colton's *The Painting Fool* or Pérez y Pérez's *Mexica* have strived to **build a brand** for their programs, by: naming their system; developing it over a long period, collating and celebrating outputs as in (Perez y Perez 2017), writing a plethora of research papers, substantial public engagement, popular press coverage, etc. The notion of an 'aura' around art works and artists is a well known concept (Benjamin 1968), and it's not impossible to imagine software having such a reputation, which could be used to add authenticity to its practice and products.

Ultimately, a lack of life experience of concepts such as love and scenarios such as childbirth leads people to projections of inauthenticity onto software when they create artefacts addressing such things. Software does, however, have life experiences, but not those that people have. For instance, The Painting Fool has interacted with and painted portraits of around 1,000 people, including a few famous people, in multiple countries. It has made people laugh, caused excitement, disappointment and interest and been written about by scores of journalists. One practical way of addressing issues of inauthenticity, is for the software to record and **use life experiences** of this nature in its creative process. That is, the software could record aspects of its creative process, outputs, public and private engagements, then refer to this data in future projects.

While it may be difficult to convince audiences to see the world from the software's perspective and that such computational life experiences are worth celebrating artistically, such an approach wouldn't suffer from being seen as inauthentic. However, the programmatic origins of the software may throw up two difficulties in the general acceptance of the notion that software has its own authentic life experiences. That is, being an engineered software entity may make it difficult for people to (a) project something akin to creative personhood onto software, and (b) empathise with something which is very different to people. It appears that we need to engineer software which moves away from the humanity of the programmer, while simultaneously moving towards the humanity of its audience members.

The *Lovelace Objection*, as framed by Turing (1950), gets to the heart of the issues surrounding the perception of authenticity, i.e., to an onlooker, the productions of a computational system may appear inauthentic because it is natural to look to the programmer as the source of authentic experiences of the world. Turing's response to the objection is remarkable for the scope of his vision of creating 'child machines' with a small set of core features, such that they can be educated (Turing 1950). Consequently, tackling issues of authenticity in autonomous creative systems may require researchers to rethink their role.

Grounding Computational Creativity

The question of authenticity is tied to the question of whether a creative system that is not *grounded* (Brooks 1991) in its world can produce anything authentic. Situated cognition argues that all knowledge is situated in activity bound to social, cultural and physical contexts, and hence that cognition is inseparable from action (Clancey 1997). Embodied cognition argues that many features of cognition, including high-level representations and reasoning (Lakoff and Johnson 1999), are shaped by aspects of the physical body of the agent (Anderson 2003). Enactivism builds on situated and embodied cognition by arguing that cognition in biological systems is not only grounded in action (Noë 2004) but is also driven by a purpose to maintain its existence as a unity (Varela, Thompson, and Rosch 1991).

AI researchers have typically adopted approaches inspired by situated and embodied cognition either to exploit the advantages inherent in embodied agents (Brooks 1991), or because embodiment is considered a necessary condition of any model of animal or human cognition (Ziemke 2004). Attempts to apply situated and embodied AI in creative systems have often focussed on advantages to be gained by developing robotic systems but also tackle questions of grounding of creativity in artificial systems.

The marimba playing robot *Shimon* (`shimonrobot.com`) is an interesting example of an embodied creative system. Internally, the medium of music is represented as a choreography of gestures, opening up new opportunities for expressive performance (Hoffman and Weinberg 2010). Another example of an embodied creative system is the painting robot *e-David* (`e-david.org`), which attempts to approximate a given photograph through an iterative process of refinement; planning, applying then reviewing the results of paint strokes (Lindemeier et al. 2015), producing an ongoing 'conversation with the medium' (Schön 1983). Such embodied creative agents avoid symbolic representations by using "the world as its own model" (Brooks 1991) and engaging in a process similar to distributed cognition (Clark 1996). The performance of these creative systems, acting and responding to their physical and social environments, are reminiscent of craftsmen in action, potentially supporting the perception that the products of the machine labour is more authentic than that of a disembodied creative system.

Intrinsic Motivation Computational models of intrinsic motivation (Oudeyer 2008) allow developers of creative systems to further distance themselves; rather than provide externally defined goals, e.g., produce works in a given style, intrinsically motivated systems are provided with inherent drives, e.g., reward signals for the discovery of novelty (Schmidhuber 1991), maximising "empowerment" (Guckelsberger, Salge, and Colton 2017), or "learning progress" (Merrick and Maher 2009). Researchers in developmental robotics (Oudeyer 2012) use computational models of intrinsic motivation to produce embodied agents able to learn how to interact with their environment (Oudeyer, Kaplan, and Hafner 2007), in line with Turing's original vision.

Forms of intrinsic motivation explored in the development of creative systems include *curiosity*, the drive to discover

novelty, and *competence*, the drive to master a skill. As an example of an embodied creative system, Merrick (2008) developed intrinsically motivated robot toys as a platform for stimulating creative play. Saunders, Chee, and Gemeinboeck (2013) developed a collective of curious robots that used their embodiment to reduce computational requirements. The process of training intrinsically motivated systems has similarities to an apprenticeship such that the creative system is guided through a sequence of learning experiences. What is not clear is how knowledge of such an apprenticeship might change the perception of authenticity.

Enactive Computational Creativity

The *enactive AI* framework (Froese and Ziemke 2009) adopts *autopoietic enactivism*, which roots intentional agency in the need of living organisms to self-produce through their ability to perceive and interact with their environment. Consequently, enactive AI extends embodied AI by grounding sensorimotor interaction in an agent's maintenance of its identity. Guckelsberger, Salge, and Colton (2017) argue that situated and embodied AI approaches do not go far enough in grounding creative systems but that *enactive AI* provides a framework for developing autonomous creative systems. From the perspective of developing autonomous creative systems, proponents of an enactive approach argue that simply being embodied is insufficient because it does not preclude the external assignment of values. For example, many of Shimon's goals are hard-coded, allowing it to improvise effectively with other musicians, while at the same time undermining its claim to autonomy because it does not act in these ways for its own purpose. Guckelsberger, Salge, and Colton (2017) propose that an enactive approach to the development of autonomous creative systems provides a method for escaping the imposition of human-given or hard-coded value systems, while at the same time conceding that such systems may not be recognised as creative due to the *embodiment distance* between the enactive system and the human observer.

Bridging the Embodiment Distance Given the challenge of adopting an enactive approach to creative systems, how might we bridge the embodiment distance between autonomous creative systems and human observers? A possible way forward is to recognise that the situation of an autonomous creative system includes the social and cultural environment, to the extent that it can share it. The embodied creative system "Curious Whispers" allows participants to interact using a three-button synthesiser, permitting the composition and performance of tunes similar to those shared between a group of robots (Saunders, Chee, and Gemeinboeck 2013). Opening up the collective to external perturbations in this way allowed some participants to inject cultural knowledge, in the form of simple tunes, into the collective memory of the agents through repeated performances. While not an enactive system, "Curious Whispers" suggests that by carefully designing creative systems to be open to their social and cultural environment in ways that allow the system to ground the incoming signals is one way that the distance between artificial and human embodiments may be bridged. Importantly, this process may open up autonomous creative systems to the social norms and cultural traditions that inform perceptions of authenticity.

Conclusions and Future Work

For culturally acceptable, truly autonomous creative behaviour in artificial systems, we believe a lack of authenticity is a looming issue. We have motivated and expanded on this belief here, and situated it in the context of *acceptable non-authenticity*, problems with a *lack of experience*, and notions of *expressive and brand authenticity*. As the quality of outputs increases, we can envisage an *uncanny valley* stretching out, where audiences marvel at the value of the products from creative systems, while despairing at the lack of authenticity in the process and in the nature of the originator. We have suggested software owning its non-authenticity, emphasising the product, producing speculative rather than realistic works of fiction, and building a brand as short-term ways in which to sidestep issues of authenticity. We have further suggested that software can record and later refer to its life experiences as a practical way in which to attain authenticity. Finally, we have discussed embodied Computational Creativity practices and proposals for enactive, purposeful computational creativity systems as ways in which we can engineer software which is simultaneously distanced from its programmers while closer to its audiences, potentially occupying an authentic position as an individual.

Addressing the (in)authenticity of software will contribute to the development of more sophisticated evaluation methods for Computational Creativity. These question the autonomy of the software, how it was constructed, what it does, how audiences and other stakeholders perceive it, how it presents its work through framing and other methods, and – as per the discussion here – should now also ask whether the software is seen as authentic in a particular project. As with these previous steps forward, we hope that acknowledging issues of inauthenticity will drive forward practical matters of engineering and deploying creative software, whether this involves simply avoiding asking software to autonomously generate emotion-laden poems about love, or developing embodied systems capable of grounding their 'life experiences' in order to authentically utilise them in future creative processes.

We hope to further highlight and unpick issues of authenticity in autonomously creative systems via the lens of existing theories on creative behaviour, such as the *Four Ps* breakdown of creativity into perspectives of person, process, product and press (Jordanous 2016). The discussion here is incomplete and too short to do justice to such a complex notion as computational authenticity, but we hope it provides a starting point for a conversation about what we believe will become an essential issue in Computational Creativity.

Acknowledgements

We wish to thank Christian Guckelsberger for his excellent and extensive comments on a draft of this paper, and the anonymous reviewers for many insightful suggestions. This work has been funded by EC FP7 grant 621403 (GRO).

References

Anderson, M. L. 2003. Embodied cognition: A field guide. *Artificial Intelligence* 149(1).

Baker, N. 2009. The Greatest Liar. https://archives.cjr.org/second_read/the_greatest_liar_1.php

Barthes, R. 1967. The death of the author. *Aspen* 5-6.

Benjamin, W. 1968. The work of art in the age of mechanical reproduction. In *Illuminations*. Shocken Books.

Brooks, R. A. 1991. Intelligence without reason. In *Proceedings of the 12th International Joint Conference on AI*.

Clancey, W. J. 1997. *Situated Cognition: On Human Knowledge and Computer Representations*. CUP.

Clark, A. 1996. *Being There: Putting Brain, Body, and World Together Again*. MIT Press.

Cohen, E. 1988. Authenticity and commoditization in tourism. *Annals of Tourism Research* 15(3).

Colton, S.; Cook, M.; Hepworth, R.; and Pease, A. 2014. On acid drops and teardrops: Observer issues in Computational Creativity. In *Proceedings of the AISB symposium on AI and Philosophy*.

Colton, S. 2012. The Painting Fool: Stories from building an automated painter. In McCormack and d'Inverno (eds) *Computers and Creativity*. Springer.

Dutton, D. 2003. Authenticity in art. In Levinson, J., ed., *The Oxford Handbook of Aesthetics*. OUP.

Elgammal, A.; Liu, B.; Elhoseiny, M.; and Mazzone, M. 2017. CAN: creative adversarial networks, generating "Art" by learning about styles and deviating from style norms. *arXiv 1706.07068*.

Errington, P. W. 2017. *J.K. Rowling: A Bibliography*. Bloomsbury.

Frey, J. 2003. *A Million Little Pieces*. Random House.

Frey, J. 2006. A note to the reader. penguinrandomhouse.biz/media/pdfs/AMLP020106.pdf

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

Galbraith, R. 2013. *The Cuckoo's Calling*. Sphere.

Gatys, L.; Ecker, A.; and Bethge, M. 2016. Image style transfer using convolutional neural networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Glaveanu, V. 2017. Creativity in craft. In *The Cambridge Handbook of Creativity Across Domains*. Cambridge University Press.

Guckelsberger, C.; Salge, C.; and Colton, S. 2017. Addressing the "why?" in Computational Creativity: A non-anthropocentric, minimal model of intentional creative agency. In *Proc. 8th ICCC*.

Hoffman, G., and Weinberg, G. 2010. Gesture-based human-robot jazz improvisation. *IEEE Int. Conf. on Robotics & Automation*.

Jacobson, D. 2018. RockNotes: Punks vs. Poseurs. beachwoodreporter.com/music/rocknotes_punks_vs_poseurs.php

Jordanous, A. 2016. Four PPPPerspectives on Computational Creativity in theory and in practice. *Connection Science* 28.

Lakoff, G., and Johnson, M. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Collection of Jamie and Michael Kassler. Basic Books.

Lindemeier, T.; Metzner, J.; Pollak, L.; and Deussen, O. 2015. Hardware-based non-photorealistic rendering using a painting robot. In *Computer Graphics Forum* 34.

Lindholm, C. 2013. The rise of expressive authenticity. *Anthropological Quarterly* 86(2).

Llano, M.; Colton, S.; Hepworth, R.; and Gow, J. 2016. Automated fictional ideation via knowledge base manipulation. *Journal of Cognitive Computation* 8(2).

MacCannell, D. 1973. Staged authenticity: Arrangements of social space in tourist settings. *The American J. of Sociology* 79(3).

Macedo, L., and Cardosa, A. 2001. Creativity and surprise. In *Proc. AISB Symp. on AI and Creativity in Arts and Science*.

Merrick, K. 2008. Designing toys that come alive: Curious robots for creative play. In *Proc. Seventh International Conference on Entertainment Computing*. Springer.

Merrick, K. E. and Maher, M. L. 2009. *Motivated Reinforcement Learning*. Springer.

Morhart, F.; Malär, L.; Guèvremont, A.; Girardin, F.; and Grohmann, B. 2015. Brand authenticity: An integrative framework and measurement scale. *J. of Consumer Psychology* 25.

Noë, A. 2004. *Action in Perception*. MIT Press.

Oudeyer, P.-Y.; Kaplan, F.; and Hafner, V. V. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation* 11(2).

Oudeyer, P.-Y. 2008. How can we define intrinsic motivation? In *Proc. 8th Int. Conf. on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. 93–101

Oudeyer, P.-Y. 2012. Developmental robotics. In Seel, N. M., ed., *Encyclopedia of the Sciences of Learning*. Springer, 969–972

Pease, A., and Colton, S. 2012. On impact and evaluation in Computational Creativity: A discussion of the Turing test and an alternative proposal. In *Proc. of the AISB symp. on AI and Philosophy*.

Pérez y Pérez, R. 2017. Mexica: 20 years – 20 stories. Counterpath.

Saunders, R., and Gero, J. S. 2001. A curious design agent: A computational model of novelty-seeking behaviour in design. In *Proceedings of CAADRIA 2001*.

Saunders, R.; Chee, E.; and Gemeinboeck, P. 2013. Evaluating human-robot interaction with embodied creative systems. In *Proceedings of the 4th Int. Conference on Computational Creativity*.

Schmidhuber, J. 1991. A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proc. of the Int. Conf. on Simulation of Adaptive Behaviour*.

Schön, D. A. 1983. *The reflective practitioner: How professionals think in action*. Basic Books.

Seyama, J., and Nagayama, R. S. 2007. The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoper. Virtual Environ.* 16(4).

Slater, L. 2004. *Opening Skinner's Box: Great Psychology Experiments of the Twentieth Century*. W. W. Norton and Company.

Turing, A. 1948. Intelligent machinery. Technical report, NPL.

Turing, A. 1950. Computing machinery & intelligence. *Mind* 49.

Varela, F. J.; Thompson, E.; and Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Varga, S. 2013. *Authenticity as an Ethical Ideal*. Routledge.

Wang, N. 1999. Rethinking authenticity in tourism experience. *Annals of Tourism Research* 26(2).

Wimsatt, W., and Beardsley, M. 1954. *The Verbal Icon: Studies in the Meaning of Poetry*. University of Kentucky Press.

Wood, A. M.; Linley, P. A.; Maltby, J.; Baliousis, M.; and Joseph, S. 2008. The authentic personality: A theoretical and empirical conceptualization and the development of the authenticity scale. *Journal of Counseling Psychology* 55(3).

Wyatt, E. 2006. Author Is Kicked Out of Oprah Winfrey's Book Club. nytimes.com/2006/01/27/books/27oprah.html

Ziemke, T. 2004. Embodied AI as science: Models of embodied cognition, embodied models of cognition, or both? *Embodied Artificial Intelligence*. Springer LNCS 3139.