

Empirically Grounding the Evaluation of Creative Systems: Incorporating Interaction Design

Oliver Bown

Design Lab, University of Sydney,
NSW, 2006, Australia
oliver.bown@sydney.edu.au

Abstract

In this paper I argue that the evaluation of artificial creative systems in the direct form currently practiced is not in itself empirically well-grounded, hindering the potential for incremental development in the field. I propose an approach to evaluation that is grounded in thinking about *interaction design*, and inspired by an anthropological understanding of human creative behaviour. This requires looking at interactions between systems and humans using a richer cultural model of creativity, and the application of empirically better-grounded methodological tools that view artificial creative systems as situated in cultural contexts. The applicability of the concepts ‘usability’ and ‘user experience’ are considered for creative systems evaluation, and existing evaluation frameworks including Colton’s creativity tripod and Ritchie’s 18 criteria are reviewed from this perspective.

Introduction: Evaluation, Creativity and Empiricism

This paper is concerned with the evaluation of creative systems, specifically in the area of artistic creativity (not to be confused with evaluation *by* creative systems). Whilst AI researchers in other application domains are able to observe and measure incremental improvements in their algorithms, computational creativity researchers are burdened by the inherent ambiguity in the field regarding whether algorithm or system X is better than algorithm or system Y. Incremental developments in the field are also relatively obscure to the outsider: the figurative artworks created by Harold Cohen’s celebrated automated artist, AARON, in the 1980s¹ look like the work of a competent and creative artist. As far as the artwork itself is concerned, this would appear to be as good as it gets – problem solved. But most in the field believe that we are only just beginning to develop good creative systems. Such appearances foster confusion about where we are at in the development of significant artistic creativity in computers, between a far-off goal on the one hand, and a solved problem on the other.

Cardoso, Veale, and Wiggins (2009) characterise the field as taking a pragmatic, demonstrative approach to compu-

tational creativity practice, “which sees the construction of working models as the most convincing way to drive home a point” (Cardoso, Veale, and Wiggins, 2009, p. 19). This tradition has kept the focus on innovation, distinguishing it from more theoretical studies of creativity. Nevertheless the discussions and demonstrations that surround such an approach depend on a firm relationship between empirical observations and what we claim about systems. Hence, understandably, a significant portion of the literature in the field focuses on the ‘necessary theoretical distraction’ of how to go about evaluating systems. Wiggins’ notion of evaluation (Wiggins, 2006), widely adopted in the field, requires that a system performs tasks in a way that would be deemed creative if performed by a human. But whilst simple to state, the task of concretely drawing such a conclusion about a given system maintains an opaque and vexing relationship to the various forms of empirical observations available to us.

In light of these issues, the purpose of this paper is to examine the empirical grounding underlying the evaluation of systems. Empirical grounding is defined as the practice of anchoring theoretical terms to scientifically measurable events, and is necessary for the “effectiveness of the application of knowledge” (Goldkuhl, 2004), that is essential for transforming discussions about system designs and methods into incremental scientific progress.

I argue that whilst the essential incompatibility between evaluation in computational creativity and the objective nature of optimisation found in AI may have been acknowledged from the outset, there remains a gap that has still not yet been plugged by a positive theory of evaluation in computational creativity. Further to this, I propose that the standard model of creativity in art, derived largely from Boden’s concepts, has not provided a suitable framework for thinking about where and how the evaluation of creativity applies in human artistic behaviour. To address this, it is proposed that a human-centred view, specifically the use of design-based approaches such as interaction design, can give computational creativity a thorough empirical grounding. An interaction design approach can be applied easily to existing work in computational creativity, viewing the understanding and measurement of system behaviours in terms of their interaction with human ‘users’. It offers a practical route to bringing a much-needed human and social dimension to studies of creative systems without rejecting aspirations to-

¹See AARON’s online biography at <http://www.usask.ca/art/digital.culture/wiebe/moving.html>.

wards autonomy in computational creativity software.

The Soft Side of Computational Creativity

The adjectives ‘hard’ and ‘soft’ have been used, controversially, to refer to different areas of scientific enquiry (as a precaution, they remain in quotes throughout this paper!). Diamond (1987) explains that some “areas are given the highly flattering name of hard science, because they use the firm evidence that controlled experiments and highly accurate measurements can provide,” whereas “soft sciences, as they’re pejoratively termed, are more difficult to study for obvious reasons... You can’t start... and stop [experiments] whenever you choose. You can’t control all the variables; perhaps you can’t control any variable. You may even find it hard to decide what a variable is” (Diamond, 1987, p. 35). Although many theoreticians such as Diamond reject the tone of the terms (here Diamond is arguing that soft sciences are in fact harder than hard sciences), the definitions given here usefully describe a continuum of what he understands as ‘degrees of operationalisation’. Whilst the terms may connote ‘tough’ and ‘weedy’ respectively, they also connote ‘rigid’ and well-defined levels of operationalisation versus more ‘flexible’ and loosely-defined levels of operationalisation. This distinction remains useful. A key point is that there are appropriate ways to deal with ‘soft’ concepts, above all of which is to acknowledge them as such in order to apply suitable methods. A popular perception is that ‘soft’ sciences harden as their theory and practice coevolve, with psychology and sociology given as typical examples (Nature, 2005). But doing quality ‘soft science’ would appear to be the first step towards this ambition. Computational creativity necessarily deals with both sorts of concepts, and researchers must therefore know how to work across this spectrum.

I discuss as an example Colton’s ‘creativity tripod’ (2008). Colton proposes to include in his formulation of evaluation a set of internal properties of systems, due to the limited information available when using only the end products of an automated creative process to evaluate that process (as advocated by Ritchie (2007)). He proposes that we look inside the system itself in order to gain a fuller description of the system’s processes along with its products, and thus make a more informed decision about the creativity of the system. This, he argues, is more in line with how we evaluate human creativity:

“A classic example... is Duchamp’s displaying of a urinal as a piece of art. In situations like these, consumers are really celebrating the creativity of the artist rather than the value of the artefact” (Colton, 2008, p. 15)

Colton suggests breaking down creativity into three components – a ‘creativity tripod’ of skill, appreciation and imagination – that can be sought in creative systems. He defines each of these as necessary conditions for the identification of creativity, and proposes that creativity evaluation could be built around an analysis of these properties. He performs such an analysis of his own systems, *HR* and *The Painting Fool*, and identifies the existence of each component in both systems (although he clarifies that they do

not occur simultaneously in the same version of the *Painting Fool* system).

In Colton’s analysis, skill, appreciation and imagination are not formalised, and are treated as intuitive ideas taken in the manner of Wiggins’ ‘creativity as recognised by a human’ criterion. Accordingly, Colton’s application of the terms is impressionistic. For example, he says of *The Painting Fool*’s imagination that “we wrote a scene generation module that uses an evolutionary approach to build scenes containing objects of a similar nature, such as city skylines and flower arrangements” (Colton, 2008, p. 21). From this, the reader has little hope of determining whether the ‘imagination’ criterion has been satisfied, let alone what the sub-criteria are for imagination. A further problem is that, in empirical terms, the expected order of knowledge discovery has clearly been put in reverse: imagination has been defined first as a kind of internal scene generation process, then implemented into the system, the conclusion being drawn that the system contains imagination. This abandons the critical step of enquiry into whether, having defined imagination as such and implemented it accordingly, this is *actually* a sufficient definition of imagination.

Under these circumstances, the concepts skill, appreciation and imagination cannot be distinguished from trivial pseudo-versions of themselves. Accordingly, reduction to triviality provides an easy rebuttal to such claims, and this has been performed by Ventura on Colton’s criteria (Ventura, 2008). Ventura presents a clearly trivial, unanimously uncreative computer program, and applies a similar analysis to that performed originally by Colton, concluding that the mock system has skill, appreciation and imagination. If Ventura’s system has these features, and they are sufficient for the attribution of creativity, then we must either accept the system as creative or reject the criteria as they currently stand.

Can such vague concepts be used at all, or should they be dropped altogether if they can’t be precisely formalised? I prefer to support both Colton’s initial premise – that an understanding of the inner workings of systems is as necessary to evaluating creativity as the outputs the system produces – and his identification of skill, appreciation and imagination as critical features of advanced creative systems. They are things that we would expect to see well implemented in our finest systems and there is nothing wrong with making this intuitive step. But unfortunately they are clumsy terms, and as Ventura’s analysis demonstrates, don’t look like hopeful performers at a formal level. In Diamond’s terms, they are far from being effectively operationalised, and they may never be operationalised, because in the process we would reasonably expect to device concepts that are far removed from folk terminology, just as physicists and neuroscientists have done.

A more generic scientific strategy for how to work with both rigid and flexible objects alike comes from the definitive hard scientist Richard Feynman (1974) who makes a simple appeal to what he describes as an unspoken law of science, “a kind of utter honesty—a kind of leaning over backwards” to face the problem of “how not to fool ourselves” (Feynman, 1974). He draws an analogy between

forms of habitual scientific practice and the famed cargo cults of the South Pacific, who carved wooden headphones and bamboo antennae in the hope of attracting cargo planes to land, imitating the troops they had seen during WWII. He calls upon scientists across disciplines to ask themselves, Am I making symbolic wooden headphones or real working headphones?

In the spirit of Feynman's call to 'utter honesty', an overlooked first step is to acknowledge that these terms, given our current knowledge, are extremely flexible and far-from-operationalised, which places very different demands on how we address and manipulate them as concepts. Their treatment is implicitly argument-based, meaning that no neat proof or direct basis in evidence is available to us. This makes for a very messy equivalence to the process of checking the steps of a proof or repeating a simulation experiment, with each step containing unknowns and vagaries: flexible rather than rigid science. Computational creativity needs to learn to work with vague concepts that are not easily subject to formal treatment.

Other examples of slips into the space of soft science that are likely to occur in computational creativity discourse include describing a system as 'doing something on its own' when discussing the autonomy of systems, but remaining imprecise about what the 'it' and the 'doing' specify (e.g., to say a program composes a piece of music 'on its own' requires quite a detailed analysis of the sequence of events leading to the specific configuration of musical content), and cases of comparing exploratory and transformational creativity in an interpretive manner (e.g., to classify any historical creative act as transformational requires the imposition of our own chosen categories onto incomplete historical data) (see Ritchie, 2006, for an interesting discussion).

For this reason 'soft sciences', such as social anthropology, subject the use of language to great scrutiny. The meaning of terms that cannot easily be made measurable or mathematically manipulable are instead treated with an acknowledgement of their fragility. As a part of their data gathering, anthropologists immerse themselves in cultural situations in order to be able to fully understand and successfully interpret what they observe. Immersion is necessary in order to expose the cultural content of these situations, which is not directly accessible through 'hard science' methods such as surveys, lab tests or recordings. For example, the difference between a twitch of the eye, a wink, a fake wink, a parodied wink, a burlesque of a parodied wink, and so on, might only be fully accessible to someone who has an intimate understanding of the sociocultural context in which the act occurs (Geertz, 1973). Misinterpretation of such acts is a clear source of error in the development of theory. In the 1980s, borrowing from philosopher Gilbert Ryle, anthropologist Clifford Geertz (Geertz, 1973) developed these practices into a method of 'thick description' that gave new impetus to, and validation of, the interpretative ('soft') side of anthropology as a science. Such thinking is more relevant to computational creativity than it may appear. The empirical material underlying Wiggins' 'creativity as recognised by a human' criterion, is in the first instance anthropological rather than psychological, revolving around interpretations

of culturally-situated human behaviour: in particular that we establish a shared understanding of what 'creative' means.

Geertz' advice on grounding methodology is that "if you want to understand what a science is, you should look in the first instance not at its theories or findings, and certainly not at what its apologists say about it; you should look at what the practitioners of it do" (Geertz, 1973, p. 5). This is a call to work the science's methods around the data and practices that are practically available. This may be helpful given what computational creativity practitioners do. Cardoso, Veale and Wiggins' characterisation of computational creativity practice as the construction of "working models as the most convincing way to drive home a point" (Cardoso, Veale, and Wiggins, 2009, p. 19), breaks down into two parts: the engineering excellence to create advanced creative systems, and the analysis of human social interaction in creative contexts that will be used to round off the argument. Thus a necessary direction for computational creativity is to fuse excellence in the 'hard science' area of algorithms and the 'soft science' of understanding human social interaction. The terms skill, appreciation and imagination are things that we should be seeking to better define through ('soft') computational creativity research, and cannot at the same time be used as the basis for a ('hard') test for creativity.

Characterising Artistic Creativity Using Generative and Adaptive Creativity

Value or utility is included in the vast majority of definitions of creativity (most notably (Boden, 1990)), and is critical to many applications of creativity research, such as improving organisational creativity and building creative cities. But non-cognitive processes such as biological evolution are also viewed as creative. Here, value cannot have the same meaning as it does in the context of human cognition-based creativity, because there is no agent to do the valuing. And yet this difference has not been explored in any depth. The application of theoretical concepts has tended to focus on Boden's (1990) two key distinctions in her analysis of creativity: between personal and historical creativity as indications of scope; and between combinatorial, exploratory and transformational creativity as forms of creative succession. From this point of view, creativity is tightly bound to individual human goals, and is primarily conceived of as a cognitive process that is used to discover new things of value.

This lack of attention to the variable nature of value in creativity causes confusion and has led to a poor empirical grounding for evaluation in computational creativity, precisely because much creativity occurs outside of the process of human creative cognition (in the narrower sense given above). A distinction based on different relations to value has not been taken up by the community. I draw on a distinction (Bown, 2012) between 'generative' and 'adaptive' creativity, and argue that this distinction clarifies and resolves the confusion about how value is manifest in the arts.

In (Bown, 2012) I propose a distinction between two forms of creativity based on their relationship to value: generative and adaptive creativity. Generative creativity is defined with a very broad scope, it occurs wherever new types

of things come into existence. It does not require cognition: non-human processes such as biological evolution are capable of creating new types of things, and, I argue, there are also examples of human activity in which things emerge ‘autopoietically’ without being planned or conceived of by individual humans. The role of generative creativity in art will be discussed below.

Generative creativity offers an expanded view of creativity in which the production of new types of thing is the sole criterion for creativity to have occurred, and the process by which those things are produced – whether by deities, human minds or autopoietic processes – is secondary. In human creativity, this liberates us from the possibly misleading premise that the ‘creative mind’ is necessary and sufficient for the ‘act of creation’. A framework that distinguishes between those entities can properly address the issue of when and how human thinking is associated with new things coming into existence.

Adaptive creativity on the other hand is that in which something is created by an intelligent agent in response to a need or opportunity. The distinguishing feature here is that of value or benefit – generative creativity is ‘value free’. In adaptive creativity, the agent doing the creation stands to benefit from the creative act: a link must exist between the creative agent and the beneficial return of the creative act in order for adaptive creativity to have occurred. Uncontroversial examples include solving everyday problems, such as using a coat-hanger to retrieve something from behind a wardrobe. Adaptive creativity is understood as requiring certain cognitive abilities such as mental representation, whereas generative creativity is completely blind, as in biological evolution.

Generative and adaptive creativity are not extremes at either ends of a continuum, but distinct and mutually exclusive categories – either there was a preceding purpose or there was not. However, the appearance of new things may be the sum of different episodes of generative and adaptive creativity.

Given these terms, I argue that the existing notion of the evaluation of creative systems is entirely – indeed inherently – geared towards adaptive creativity, and is unable to accommodate generative creativity at all. Adaptive creativity alone is compatible with computational creativity’s AI legacy, which preferences an optimisation or search approach to discovering valuable artefacts. This is not without powerful applications. Evolutionary optimisation regularly discovers surprising designs in response to engineering problems. Thaler’s “Creativity Machine”, for example, was used to discover novel toothbrush designs using a relatively traditional optimisation approach involving a clear objective function (Plotkin, 2009). It is only generative creativity that is incompatible with optimisation.

Adaptive and Generative Creativity in the Arts

For the purpose of evaluating creative systems, it has been considered reasonable to assume that we can treat artistic domains entirely in terms of adaptive creativity, and that the act of creating artworks is an adaptively creative act. Accordingly one can view the production of an artwork as an

optimisation or search problem. This simplification is built in to the premise of an agent designed to evaluate its output in order to find good solutions. For such an agent to incorporate generative creativity into its behaviour would mean that the value of its output was indeterminate and evaluation would be frustrated.

But evidence suggests that this view of art does not hold when one considers its social functions. I will focus on music for the purpose of this discussion, and take what I believe is an uncontroversial understanding of music insofar as sociologists of music are concerned. Hargreaves and North (1999) identify three principal social functions for music: self-identity, interpersonal relationships and mood. These in turn, they argue, shape musical preference and practice. For example, “research on the sociocultural functions of music suggests that it provides a means of defining ethnic identity” (Hargreaves and North, 1999, p. 79).

The evidence they gather shows the perceived aesthetic value of music not to be determined purely by exposure to a corpus or ‘inspiring set’, but also by a set of existing social relationships. More recent research in experimental psychology reveals an increasingly complex story behind how we give value to creative artefacts. Salganik, Dodds, and Watts (2006), for example, show that music ratings are directly influenced by one’s perception of how others rated the music, not just in the long term but at the moment of making the evaluation. Newman and Bloom (2012) examine the underlying causes of the attachment of value to originals rather than copies, finding, amongst other things, that the value given to an original is associated with its physical contact with the artist. Both studies suggest a form of winner-takes-all process whereby success begets further success. Such phenomena place limits on the importance of the creative content in evaluation. Admittedly artistic success is not the same as artistic creativity, but the overlap is great enough, in any practical sense of evaluating creativity, to carry the argument from one domain to the other.

Csikszentmihalyi’s (1999) domain-individual-field theory has long held that individuals influence domains and alter fields, but such observations have on the whole been only been acknowledged, not actually applied in computational creativity. Coming close, Charnley, Pease, and Colton (2012) present ‘framing’ as a way to deal with the process of adding additional information that may influence the value of a creative output. According to the idea of framing, I might provide information alongside an artwork, such as an exhibition catalogue entry, that influences its perception. In its simple form framing would embellish an artwork, perhaps explaining some hidden symbolism behind the materials used. But in this sense it is simply a part of the system output along with the artwork. By comparison, verbal statements, and other social actions, can have effects with respect to value that are categorically different from this, for example by provoking people to alter their perception of value in general. Framing takes steps towards the idea that value can be manipulated, even ‘created’, but continues to assume a fixed frame of reference.

Taking these additional processes into account, when an individual produces an artwork, some amount of the value

of that artwork may have already been determined by factors that are not controlled by the individual, or be later determined by factors that are unrelated to the content of the work. The creativity invested in the creation is not entirely the product of the individual, whose artistic behaviour may be more associated with habit and enculturation than discovery, but is imposed upon the individual through their context and life history. The anthropological notion of the 'dividual', or 'porous subject' (Smith, 2012) has been used to capture this idea of a person as being composed of cultural influences, indicating their ongoing permeability to influence. According to this view, the flux of influence between individuals may have an equivalence to the interaction between submodules within a single brain, meaning that isolating individuals as units of study is no better a division than focusing on couples, tuples, larger groups or cognitive submodules. Given this understanding of individual human behaviour in relation to culture in general, and the arts in particular, computational creativity can be seen to place too much emphasis on the idea of individuals being independent creators.

From this alternative point of view it is argued that artistic behaviour has a significant generative creativity element by which new forms 'spring up', not because individuals think of them, but through a jumble of social interaction. Such emergent forms may have structural properties related to the process that produced them, but they were not made with purpose. By analogy, consider a classic debate about adaptationism and form in evolutionary theory: the shape of a snail shell, as described in Thompson's *On Growth and Form* (Thompson, 1992) comes about through the process of evolutionary adaptation. But this is not purely a product of the selective pressures acting on the species. It results from an interaction between selective pressures and naturally-occurring structure. Likewise, human acts of creation are constrained by structural factors that guide the creator, augmenting agency.

The notion that a system possesses a level of creativity is riddled with complexity, owing to the fact that creativity is as much something that is enacted upon individual systems as enacted by them. In computational creativity, this means that the goal of evaluating virtual autonomous artists is not empirically well-grounded when performed in isolation. Empirical grounding requires a strong coherence between our theories and practices, and the things we can observe. In the following section, I will argue that an interaction design approach delivers this coherence, bringing together system development with a thorough understanding of the culturally-situated human. I will suggest that interaction design shouldn't be viewed merely as an add-on or a form of research used only at the application stage, but that it has a central role to play in improving methodology in computational creativity.

Towards Empirical Grounding

To reiterate the argument so far, empirical grounding is defined as the process of anchoring theoretical terms to scientifically measurable events. Computational creativity characteristically employs a makers' approach to innovating new

ideas and building better systems, but the idea of asking how creative these systems are is not empirically well-grounded. Then what can we ask? I have examined the need simply to elaborate on terms and concepts during the process of evaluation, adopting appropriate 'soft science' ways of thinking alongside the existing engineering mindset, but although a well-grounded approach needs to take this into account, it does not provide a grounding itself.

Two research methodologies already well integrated into computational creativity offer a basis for empirically well-grounded research. These are interaction design and multi-agent systems modelling. In both cases the imbalance between generative creativity and adaptive creativity is addressed. In the interaction design approach, creative systems are treated as objects that are inevitably situated in interaction with humans. The nature of that interaction, including its efficacy, is treated as the primary concern. Here the empirical grounding comes from the fact that properties of interaction and experience related to the analysis of usability and user experience can be observed and measured, whilst existing notions of creativity evaluation can easily be incorporated into theories of interaction design. This need not be limited to a creative professional working with a piece of creative software, but could apply to any form of interaction between person and creative system. In the modelling approach, artificial creative systems are treated as models of human creative systems. For the reasons discussed above, it does not suffice to test the success of model systems by attempting to evaluate their output, but many other observable and measurable aspects of human creativity can be studied. Multi-agent models of social networks are particularly appealing in this regard because generatively creative processes fall inside the scope of the system being studied, alleviating the tension between adaptive and generative creativity.

In this paper I only elaborate on the interaction design approach, firstly because it is more immediately applicable to computational creativity practice, and secondly because much of what can be said about empirically grounded modelling is well-known to researchers.

Interaction Design

Discussions of humans evaluating machines are commonplace in the computational creativity literature. But a lot less attention is paid to the wider range of ways in which humans can interact with creative systems. The word 'interaction', applied in the context of humans interacting with creative systems, was only used in three out of 41 papers in the 2013 ICCV proceedings (and six papers out of 46 in 2012).

Interaction design is a large field of research and is not presented in any depth here (a good introduction is the textbook by Rogers, Preece, and Sharp (2007)). The following discussion considers computational creativity in light of some core topics from the field, and looks beyond to how a study of interaction in its widest sense could be usefully applied to computational creativity.

A number of computational creativity studies are already explicitly user-focused owing to their specific research goals. For example, DiPaola et al. (2013) examined the use

of evolutionary design software in the hands of professional designers, looking at usability through the integration with the creative process, and ultimate creative productivity.

A human-centred approach to the evaluation of creative systems shifts the nature of the enquiry very slightly, by asking not how creative a system is, or whether it is creative by some measure, but how its creative potential is practically manifest in interactions with people. However, this does not require researchers to repurpose their systems as tools for artists, designers or end users, or abandon the goal of automating creativity, but to take a pluralistic approach to the application of creativity as something that is realised through interaction. As addressed in the work of DiPaola et al. (2013), described above, an obvious instance is to look at usability in the case of creativity support tools.

This is the classical locus of interaction between interaction design and computational creativity. But even researchers working towards fully autonomous ‘artificial artists’ are building systems that will ultimately interact with people, albeit in non-standard ways. Examples include artists such as Paul Brown (Brown, 2009), who has wrestled with the notion of maximising the agency of a system to the exclusion of the human artists’ signature. As the discussions surrounding such system design shows, there is no shortage of interaction between systems and the social worlds they inhabit, any of which can be considered a source of rich data.

Beyond usability a key concept in interaction design is ‘user experience’ (Hassenzahl and Tractinsky, 2006). User experience looks beyond efficacy with respect to function to consider a host of subjective qualities to do with interaction more generally, such as desirability, credibility, satisfaction, accessibility, boredom and so on (Rogers, Preece, and Sharp, 2007). Analysis of user experience includes understanding users’ desires, expectations and assumptions, and their overall conceptual model of the system. These diverse and quite vague concepts in user experience are arguably of greater importance than usability in a wide number of circumstances, and can also be at odds with it. For example, in game development pleasure can be seen to be contrary to usability (Rogers, Preece, and Sharp, 2007): dysfunctional ways of doing things, as embodied in interface design choices, may be more fun than more functional choices. By comparison, computational creativity need not be reduced to issues of function.

Such analytical concepts present a striking match with the most ambitious goals of computational creativity. Returning to Wiggins’ definition, it would not be surprising to find that a human’s appraisal of machine creativity is subject to a complex of user-experience design factors. Concepts such as surprise are already established in computational creativity theory, whereas other notions, such as the role of music and art in the development of social identification, are not, but may form part of the design of a successful ‘computational creativity’ experience.

To acknowledge and make explicit the design component in creating autonomous systems may help remove the perceived paradox that the system is an autonomous agent supposedly independent of its creators, by examining what ‘designed autonomy’ would actually mean.

Often successful computationally creative systems involve some kind of puppetry, such as the subtleties of ‘fine tuning’ described by Colton, Pease, and Ritchie (2001). Many working in this area have embraced the idea of creative software either as a tool, a collaborator that is not capable of full autonomy, or as a creative *domain* in its own right. In these cases the interaction between human artist and software agent is treated as a persevering and explicitly acknowledged state of affairs, rather than as a temporary stop on the way to fully autonomous creative systems.

In such cases it is again fruitful to think of the relationship between the developer/artist and the system in terms of usability, even if the working interface is simply a programming environment. Such a view may lead to better knowledge about effective development practices that in turn speed up the creation of more impressive creative systems. Accepting the role of developers and artists also enables a better grasp of the attribution of authorship and agency, asking instead a question of degree – how much and in what way the system contributed to the creative outputs – rendering unimportant the ideal of ‘full autonomy’.

From Evaluating System Creativity to Analysing Situated Creativity

Taking an interaction design approach reveals a wealth of empirically grounded questions that can be asked about creative systems without changing the basic designs and objectives of practitioners, and without an overly narrow focus on the question of how creative the system is.

But in order not to throw out the baby with the bathwater, since our interest is in systems that act creatively, then the creativity of systems must remain the focus of an interaction design approach. We require enriched ways to question the nature of creative efficacy and creative agency in systems. For example, an interaction design approach can better frame our evaluation of the issue of the software’s autonomy, which might otherwise be occluded.

A number of existing approaches to evaluation already give ample space for domain-specific and application-dependent variation in their use, but do not go so far as to preference design and interaction studies over direct evaluation in computational creativity. Jordanous’ (2011) proposal for creativity evaluation measures that are domain-specific suggests a design approach which is targeted at specific user-groups and specific needs, rather than an objective notion of what creativity is. A number of other researchers have proposed objective or semi-objective (depending on human responses) measures that are associated with creativity (they are not necessarily measures *of* creativity). Kowaliw, Dorin, and McCormack (2012), for example, compare formal definitions of creativity, written into a system, with human evaluations, so as to examine the accuracy of these definitions.

One of the most widely applied and discussed examples is Ritchie’s (2001; 2007) set of criteria. Ritchie proposes 18 criteria for “attributing creativity to a computer program”. The criteria derive from two core pieces of information that apply wherever a machine produces creative outputs: the inspiring set *I* (the input to the system) and the system’s output *R*. An evaluation scheme (often multi-person surveys

in the implementations examined by Ritchie) is then used to form two key measures for each output in *R*: *typicality* is a measure of how typical the output is of the kind of artefact being produced; *quality* is a measure of the perceived or otherwise computed quality of that artefact. From these scores, Ritchie organises the outputs into sets according to whether they fall into given ranges of typicality and quality. These sets are then applied in various ways in the calculation of the resulting Boolean criteria. For example, criterion 5 states that the number of outputs that are both high-quality and typical, divided by the number of outputs that are just typical, is greater than some given threshold (this, plus the thresholds required to determine the ‘high-quality’ and ‘typical’ sets, are left to the implementer to specify). As with all of Ritchie’s criteria, criterion 5 corresponds to a natural usage of the term creativity, in this case that a system whose set of typical outputs rarely includes valuable outputs is in some sense creatively lacking.

One practical problem with Ritchie’s criteria, as illustrated by the examples of their application to creative systems reported in (Ritchie, 2007), is the difficulty with which implementers establish their evaluation scheme. For example, Pereira et al. (2005) measure typicality based on closeness to *I*, calculated using edit distance. The appropriateness of this choice is hard to determine. Others use human responses to surveys, providing a form of empirical grounding. But such surveys may have wide variance, and the formulations of the criteria have no way of incorporating variance, which would represent a more complex model of the social system in which the creative agent operates. This belies the fact that typicality is a slippery, ‘soft science’ concept in reality and its relationship to a measure of quality more so, despite the clarity of Ritchie’s mathematics. Thus, as with Colton’s tripod, Ventura (2008) points to shortcomings in the criteria by showing that trivial programs can reveal instances of inherent insufficiency in their outcomes when compared with intuitive analysis of the same systems. The underlying problem is that of how to empirically ground the choice of evaluation scheme itself, such that it might provide an empirical grounding for the criteria, suggesting that the mathematics has simply shifted the hard problem from one place to another. The best we can do is to see how the various evaluation schemes and criteria relate in practice to other observables, thus the critical point: using human responses about creativity or related features of a system, alone, does not itself provide an empirical grounding for understanding the system, but rather a data point about the wider interaction. Further studies of behaviour are required to empirically ground our understanding of what these human responses mean.

A related issue in the discussion surrounding Ritchie’s criteria, is what to do with the results obtained. The criteria have, in Ritchie’s view, often been misunderstood as some sort of multivariate test for creativity. Confusingly, Ritchie unintentionally encourages this misunderstanding in his description of them as “criteria for attributing creativity to a computer program” (Ritchie, 2007). In fact he cautions against their direct use in this way.

Thus the criteria offer different analytical windows onto

the creative nature of systems. We are invited to preference some criteria over others, but given no advice on how to. However, from the point of view of interaction design, such ambiguity is expected and desirable. In application, we may value systems that are good at producing a high ratio of quality to overall output, or typicality to overall output, or quality within the typical set. Alternatively, other approaches to creativity may suggest counter-intuitive additions or alterations to Ritchie’s criteria, such as novelty search (e.g., Lehman and Stanley, 2011), which attempts to chart an output space by relentlessly searching for atypicality. The result of this is a broad representative spread of prototypes, not a concentration of high-value or typical outputs, so would score low on many of Ritchie’s criteria but may prove to be the basis for powerful automated creativity. An interaction design approach is implicit in Ritchie’s treatment of systems as tools. For example, in defining typicality, he refers to the system as having a job to do “producing artefacts of the required sort” (Ritchie, 2007, p. 73). This is not, on reflection, a requirement associated with being creative, but with performing some function required by the user or designer.

With this in mind, is it possible that the final step of “attributing creativity to a computer program” has caused more confusion than clarity, and should be quietly dropped? I suggest that it should and, echoing Jordanous (2011), that the criteria are better suited to specific creative scenarios. A jingle composer may preference typicality and require only an average degree of value, whereas an experimental artist may have little or no interest in typicality but is willing to hold out for rare instances of exceptional value. Both have different time-demands, resources, goals, aesthetic preferences and notions of the role of creativity in their work. It would not be unusual to view the experimental artist as the more creative of the two, but this is clearly only an assumption given our present theoretical understanding of creativity. The same applies to end users. Even a consumer may want typicality sometimes, and extraordinary experiences at other times.

Thus the problems raised concerning Ritchie’s criteria and their application are very easily addressed by taking a human-centred view of creative systems. Applications in the domain of both generative and adaptive creativity can be devised, and examination of the creative behaviour of systems can then be empirically well-grounded in the methods of interaction design.

Conclusion

The main argument of this paper is that the evaluation of systems as it is currently typically conceived in the computational creativity literature is not in itself empirically well-grounded. The data provided by performing human evaluations should instead be understood as one potential source of information that can feed into studies of the interaction between creative systems and people in order to be well-grounded. Systems may only be understood as creative by looking at their interaction with humans using appropriate methodological tools. A suitable methodology would include, (i) the recognition and rigorous application of ‘soft science’ methods wherever vague unoperationalised terms

and interpretative language is used, and (ii) an appropriate model of creativity in culture and art that includes the recognition of humans as ‘porous subjects’, and the significant role played by generative creativity in the dynamics of artistic behaviour. For the time being at least, terms such as ‘creativity’ and ‘imagination’ do not describe things that we can readily measure or objectively identify, they are concepts that frame other kinds of measurable and objectively identifiable things, as part of a loose theoretical framework.

References

- Boden, M. 1990. *The Creative Mind*. George Weidenfeld and Nicholson Ltd.
- Bown, O. 2012. Generative and adaptive creativity. In McCormack, J., and d’Inverno, M., eds., *Computers and Creativity*. Berlin: Springer. 361–381.
- Brown, P. 2009. Autonomy, signature and creativity. In McCormack, J., and d’Inverno, M., eds., *Dagstuhl Seminar Proceedings 09291: Computational Creativity: An Interdisciplinary Approach*, 1–7.
- Cardoso, A.; Veale, T.; and Wiggins, G. A. 2009. Converging on the divergent: The history (and future) of the international joint workshops in computational creativity. *AI Magazine* 30(3):15.
- Charnley, J.; Pease, A.; and Colton, S. 2012. On the notion of framing in computational creativity. In *Proceedings of the Third International Conference on Computational Creativity*, 77–82.
- Colton, S.; Pease, A.; and Ritchie, G. 2001. The effect of input knowledge on creativity. In *Case-based reasoning: Papers from the workshop programme at ICCBR*, volume 1.
- Colton, S. 2008. Creativity versus the perception of creativity in computational systems. In *AAAI Spring Symposium: Creative Intelligent Systems*, 14–20.
- Csikszentmihalyi, M. 1999. Implications of a systems perspective for the study of creativity. In Sternberg, R. J., ed., *The Handbook of Creativity*. New York: Cambridge University Press. 313–335.
- Diamond, J. 1987. Soft sciences are often harder than hard sciences. *Discover* 8(8):34–39.
- DiPaola, S.; McCaig, G.; Carlson, K.; Salevati, S.; and Sorenson, N. 2013. Adaptation of an autonomous creative evolutionary system for real-world design application based on creative cognition. In *Proceedings of the Fourth International Conference on Computational Creativity*, 40–48.
- Feynman, R. 1974. Cargo cult science. Available from http://neurotheory.columbia.edu/ken/cargo_cult.html.
- Geertz, C. 1973. *The Interpretation of Cultures*. New York: Basic Books.
- Goldkuhl, G. 2004. Design theories in information systems—a need for multi-grounding. *Journal of Information Technology Theory and Application (JITTA)* 6(2):7.
- Hargreaves, D. J., and North, A. C. 1999. The functions of music in everyday life: Redefining the social in music psychology. *Psychology of Music* 27(1):71–83.
- Hassenzahl, M., and Tractinsky, N. 2006. User experience—a research agenda. *Behaviour & Information Technology* 25(2):91–97.
- Jordanous, A. 2011. Evaluating evaluation: Assessing progress in computational creativity research. In *Proceedings of the second international conference on computational creativity (ICCC-11)*. Mexico City, Mexico, 102–107.
- Kowaliw, T.; Dorin, A.; and McCormack, J. 2012. Promoting creative design in interactive evolutionary computation. *IEEE transactions on evolutionary computation* 16(4):523.
- Lehman, J., and Stanley, K. 2011. Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation* 19(2):189–223.
- Nature. 2005. In praise of soft science. *Nature* 6(23):1003.
- Newman, G. E., and Bloom, P. 2012. Art and authenticity: The importance of originals in judgments of value. *Journal of Experimental Psychology: General* 141(3):558.
- Pereira, F. C.; Mendes, M.; Gervás, P.; and Cardoso, A. 2005. Experiments with assessment of creative systems: an application of Ritchie’s criteria. In *Proceedings of the workshop on computational creativity, 19th international joint conference on artificial intelligence*.
- Plotkin, R. 2009. *The genie in the machine: how computer-automated inventing is revolutionizing law and business*. Stanford University Press.
- Ritchie, G. 2001. Assessing creativity. In Wiggins, G. A., ed., *Proc. of AISB’01 Symposium*.
- Ritchie, G. 2006. The transformational creativity hypothesis. *New Generation Computing* 24(3):241–266.
- Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17(1):67–99.
- Rogers, Y.; Preece, J.; and Sharp, H. 2007. Interaction design.
- Salganik, M. J.; Dodds, P. S.; and Watts, D. J. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *science* 311(5762):854–856.
- Smith, K. 2012. From dividual and individual selves to porous subjects. *The Australian Journal of Anthropology* 23(1):50–64.
- Thompson, D. W. 1992. *On Growth and Form*. Dover.
- Ventura, D. 2008. A reductio ad absurdum experiment in sufficiency for evaluating (computational) creative systems. In *Proceedings of the 5th international joint workshop on computational creativity*. Madrid, Spain, 11–19.
- Wiggins, G. A. 2006. Searching for computational creativity. *New Generation Computing* 24(3):209–222.