

Computational Artistic Creativity and its Evaluation

David C. Brown

Computer Science Department, WPI, Worcester, MA 01609, USA

“For these [art] students, *creativity* was a ‘lovey-dovey cliché used by people who are not professionally involved with art.’” (Thornton 2008).

1. Introduction

This seminar focuses on the possibility of computational production of “artistic creativity”. That is, the production by a computer of some artifact that people would be prepared to label as a type of “art” (e.g., a painting, a piece of music or a piece of writing), where that artifact can be judged as creative.

As the existence of creativity is a judgment, relative to personal or group norms (Boden 1994) (Amabile 1993), it is important to understand what factors affect that judgment. Hence the “seminar question”— what are the appropriate methods and measures to objectively verify and validate creative behavior in artificial systems?

In most cases, people judge creativity by evaluating the resulting artifact, after the productive behavior has been completed. Hence, we will concentrate on the artifact here. However, sometimes the artistic process is inferable by the evaluator, perhaps due to some attributes of the artifact, or sometimes it’s even observable. Colton (2008) argues that art is identified as more creative if the process that generated it is considered to be more creative.

As processes that can be judged as creative can lead to very dull results, and dull processes can lead to artifacts that can be judged as creative, it’s best for the moment to focus on just one of the two (the artifact), and avoid these complex interactions.

Creativity is an intelligent activity: intelligent activity is an area studied by AI. Hence computational creativity usually incorporates AI techniques. However, a significant issue for all AI research is the *goal* of the research. The goal of a researcher affects the methods they use, the assumptions they make, as well as the criteria used to determine the success or failure of their work. This is true too of computational creativity research. Consequently to compare, contrast and evaluate such research, the goals must be made clear.

Russell and Norvig (2003), for example, describe four different approaches to studying AI. One for example is concerned with simulating human thinking (“the cognitive modeling approach”) while another with acting rationally (“the rational agent approach”). These are analogous to focusing on producing creative reasoning versus producing creative results.

The distinction between focusing on obtaining creative results, perhaps even regardless of how they are obtained, as opposed to making hypotheses about and simulating the underlying processes, is important. Systems based on Genetic Programming, for example, are able to produce remarkable results (Spector 2008). As they use a general method, it appears unlikely that they will tell us much about creativity: apart from the fact that computers are able to produce artifacts that are judged as creative. But that we already know.

To my mind the two key ways to benefit from the computational study of creativity are a) to study what kinds of knowledge and reasoning might be part of a process that produces artifacts that are judged to be creative, and b) to study the criteria that a person or a system might use to judge whether an item is creative or not.

In an earlier paper (Brown 2008) we made a similar point regarding Computational Design Creativity: i.e., the design of products, rather than artistic artifacts. In this paper we will focus on “b”, evaluation criteria. There already exists at least one well-tested set of attributes that people use to evaluate products for creativity (Besemer 2006). Products can also be evaluated by how well they meet their Requirements.

For artistic creativity, requirements may not exist, and constraints on the artifact (the artistic “product”) are usually looser or absent. Many computational creativity systems produce artistic artifacts, but such results can be judged in a variety of ways: by a variety of artistic standards or by the perceiver’s “taste”, for example. There is less chance of a generated artifact being judged in a single, clear and concrete fashion, so the standards may be softer and perhaps easier to satisfy. With regard to taste, Boden (1994) quotes, “I don’t know anything about art, but I know what I like”. If this were true in general, then there would be as many tests of the creativity of an artifact as there are people!

2. Evaluation Criteria

Evaluation is needed during the creation process to prune possibilities in the very large spaces of possible results that exist with both products and artistic artifacts. It seems to be the case that a system or an individual can’t deliberately be creative without the ability to evaluate some or all of the factors that make the result creative (Boden 1994) (Eysenck 1994) (Minsky 2006).

For products, creativity assessment includes some evaluation of quality. Not all products are intended to be creative. For art, the underlying intention is already to be creative in some way or another. This author is not aware of any literature that lists and describes

precise criteria used to evaluate how creative an artistic artifact might be: however, there is a lot of work discussing aesthetic properties.

Besemer's Creative Product Analysis Model (CPAM) (Besemer & Treffinger 1981) is the basis for a well-established, well-validated and practical product creativity assessment instrument called CPSS (Besemer 2006) (Horn & Salvendy 2006) (O'Quin & Besemer 1989). The model has three main factors: *Novelty*, *Resolution* and *Style*. Each of these factors has between 2-4 facets that further refine them. The CPSS is currently a web-based collection of questions that produces scores for each of the nine facets. The shape of the resulting creativity histogram indicates variations in creativity. The histogram acts as a creativity profile for the product. The factors and their facets are as follows (Besemer 2006).

Novelty is “the extent of newness in the product” and refers to the “number and extent” of the new processes, new techniques and new concepts included in the product. It also refers to “the newness of the product both in and out of the field”. This factor is common to most models of creativity. The facets of novelty are:

Surprising: “The product presents unexpected or unanticipated information to the user, listener, or viewer”.

Original: “The product is unusual or infrequently seen in the universe of products made by people with similar experience and training”.

Resolution is “the degree to which the product fits or meets the needs of the problematic situation”. For products at least, new but bizarre objects aren't seen as creative, as products are usually associated with an intended function, and therefore being ‘useful’ is prized. The facets of resolution are:

Logical: “The product or solution follows the acceptable and understood rules for the discipline”.

Useful: “The product has clear practical applications”.

Valuable: “The product is judged worthy because it fills a financial, physical, social, or psychological need”.

Understandable: “The product is presented in a communicative, self-disclosing way, which is ‘user-friendly’”.

Style is “the degree to which the product combines unlike elements into a refined, developed, coherent whole, statement or unit”; “how the product presents itself”; the “product's personality”. It affects how creative the product is perceived to be, and may even impact how novel it seems. For example, a telephone covered in whittled wood would create a very different impression than a phone covered in fine leather. The facets of style are:

Organic: “The product has a sense of wholeness or completeness about it. All the parts work well together”.

Well-Crafted: “The product has been worked and reworked with care to develop it to its highest possible level for this point in time”.

Elegant: “The product shows a solution that is expressed in a refined, understated way”.

Jirousek (1995) writes about the evaluation of visual design for functional objects: i.e., decorative objects with an intended purpose. Note that she is not explicitly concerned with evaluating creativity, just the quality of the visual components of the design. As her specialty is textiles, she approaches her criteria with a strong artistic bias, as opposed to Besemer's purely product bias. She provides the distinguishing example of a bra made of barbed wire and plumbing fixtures—clearly not intended to function as clothing, but as art. We will return to the function of art later. She distinguishes between the essential structure, and the “applied decorative design”. The former affects function, while the latter can but need not. Her criteria are as follows (Jirousek 1995).

Form/Function: “Does the form follow function? Is the form or shape suitable to the intended purpose of the object?”

Appropriateness of Materials: “Does the material used suit the function? Are the materials used appropriate and practical for the purpose?”

Honest Use of Materials: “a) Are the materials used honestly? Does the material look like itself, or is it disguised? b) If the form and materials are disguised, is it justified? If the form and material are used imitatively, does the imitation make a positive aesthetic contribution?”

Purpose: “What is the purpose of the applied decorative design? Is it necessary? Does it enhance the object's function?”

Concept: “Is the applied decorative design harmonious in design and idea with the purpose of the object?”

Spatial Placement: “Is the design ordered to relate to the space on which it is placed?”

Structural Coherence: “Does the decorative design reinforce, or hide the structure?”

Aesthetics: “Is the form and decoration well executed in terms of style, skill, and craftsmanship?”

Teleis: “a) Are the form, decoration, and materials expressive of a particular time and culture? b) If they are expressive of a particular time and culture, are they utilized in a context that is congruent with and respectful of their origins?”

Innovation: “Is the design a new expression, or merely a superficial change? Does it have individuality, or is it a cliché?”

Longevity: “Has the form and design an appeal that can be sustained during the length of time the object will be used? Is it destined to be a fad or a classic?”

Much more generally, Abrams (1986) describes four ways of considering the analysis, criticism or evaluation of art: Mimetic; Pragmatic; Expressive; Objective. *Mimetic* is concerned with imitating aspects of the observable universe, other than the artifact, artist or audience (i.e., artist as mirror). *Pragmatic* is concerned with the relationship between the artifact and the audience. This includes teaching and producing emotional reactions. *Expressive* is concerned with externalizing their inner life (i.e., artist as lamp). *Objective* is concerned with viewing the artifact in isolation. Of course, combinations are possible.

3. Carrying Out an Evaluation

One problem with art is that it “accumulates meaning through an extended collaborative act”, and that “Great art is essentially work that has proven inexhaustible in terms of the

value it gives to those who pay attention to it”: that is, its value is in terms of its accumulated “communicative connotations” (Thornton 2008). It is unclear whether a *single* source of evaluation is sufficient. But let’s suspend disbelief...

Based on the selected evaluation criteria presented above there are a variety of issues to address.

The first is whether Besemer’s nine facets (2006) *can* be used to evaluate the creativity of an artistic artifact, and also whether they *should*. The facets were developed after comprehensive and careful statistical analysis of terms used by human subjects to label products as creative or not. As they were judging ‘products’ with intended functionality, where practical use was fundamental, the application of Besemer’s nine facets to art is suspect, at best. However, it should be educational to try. The facets are Surprising, Original; Logical, Useful, Valuable, Understandable; Organic, Well-Crafted, Elegant.

Art is almost by definition “original”, and is often “surprising”, especially for modern art. Note that the time of evaluation (or ‘viewing’ for art) is key: looking at an impressionist painting for the first time from today’s context, is quite different from seeing it for the first time at the time it was painted. It changes the possible judgments that a viewer can make. “Logical” is about following the rules for the discipline, and hence this will vary enormously with the type of media, and the intent of the artist: some art is about breaking the rules in order to increase originality. Art is not “useful” in the same way as a product is, and, although it may have a purpose, that is often unclear. Art can indeed be “valuable”, both in financial terms, and in other ways, so this facet seems to be a relevant. As for “understandable”, it is unusual for art to be “self-disclosing” and “user-friendly”. It is possible for an artistic artifact to be “organic” but not necessary. While art may be “well-crafted”, in the sense of “reworked”, that isn’t always clear to the viewer. In addition, it isn’t necessary for an artistic artifact to be “elegant” in order for it to be seen as creative or have high quality. So, overall, the result seems to be a partial match, but with most being very hard to produce computationally.

We should expect Jirousek’s (1995) criteria for the evaluation of visual design for functional objects to be more successful than Besemer’s when applied to art, as the former is more concerned with artistic aspects.

If we knew what the function of the artistic artifact actually was, then a judgment might be made about the relationship between form and function. Material suitability also depends on function, so that’s hard to judge. The “honest” use of materials might be possible to judge for an artwork, although it raises the problem of assessing a “positive aesthetic contribution”, which is rather recursive. Harmony and necessity relate to purpose. Can we even ask whether an artistic artifact is “necessary”? As for the purpose of the “applied decorative design”, we’ll discuss this below. “Spatial Placement” and “Structural Coherence” both relate to the decorative design on a functional object, and, as a consequence are hard to apply to an artwork. The evaluation of “aesthetics” “in terms of style, skill, and craftsmanship” should be applicable to *some* art, but may require insight into the process used to produce the artifact. The issue of whether “form, decoration, and

materials” are “expressive of a particular time and culture” should apply to art, but is clearly non-trivial to assess. Innovation in art is similar to that in products: important for creativity but hard to judge. As for “longevity”, if dealers knew whether an artifact was to be a “fad or a classic” then they’d be very successful. However, making that judgment actually implies judging the quality of the artwork overall, plus understanding cultural trends: effectively predicting the future.

One thing in common between Besemer’s and Jirousek’s systems of evaluation is that they both address the “purpose” of an artifact: i.e., what the artifact is for; why it exists. If these systems are to apply to art, then we need to ask “what is the purpose of art?” or more specifically, “what is the purpose of this particular artistic artifact?”.

Jirousek (1995) suggests the following list of the purposes of art:

- A vehicle for religious ritual;
- Commemoration of an important event;
- Propaganda;
- Social Commentary;
- Recording of visual data;
- Creating beauty;
- Story telling;
- Conveying intense emotion.

An alternative, overlapping list (Wikipedia 2009) suggests that art is for:

- Communication;
- Entertainment;
- Political change;
- Psychological and healing purposes;
- Social inquiry, subversion and/or anarchy;
- Propaganda or commercialism.

Could any of these be detected reliably by a person? Could any of these be detected reliably by a computational system? It seems unlikely: but for any evaluation criteria that relate to purpose we would need to. As the titles of artistic artifacts become more obscure (e.g., *The Physical Impossibility of Death in the Mind of Someone Living*) and provide less assistance, the viewer has to decide which possible purpose or purposes the work might have, and pick the most likely.

In a similar way, the evaluation of an artistic artifact might depend on whether it was possible to detect whether the artist intended a mimetic, pragmatic, expressive or objective stance (Abrams 1986), as well as which stance the viewer adopts. Evaluating its imitative qualities, its ability to teach or produce emotions, its ability to convey the artists “inner life”, all seem very hard. In addition, evaluating the artifact as an artifact (art for art’s sake) requires knowledge about art, its conventions and forms.

4. Conclusions

It is clear that the quality and creativity of an artistic artifact is evaluated differently depending on whom is doing the evaluation: e.g., “beauty is in the eye of the beholder”. It also depends on the time, as well as physical and cultural context of the evaluation. Some past art was intended to show beautiful things, and past evaluations were done with that expectation. These days artists struggle to express themselves, try to affect our emotions or teach us about extreme social conditions. Art in general is becoming more extreme.

Our evaluations of an artistic artifact can be done in a very large number of different ways. How an artwork affects our emotions will vary from person to person. Whether we can detect the purpose of an artwork and think it worthy will vary from person to person. Whether emotions or purpose can be determined by a computer is in great doubt. Even the fundamental judgment of novelty is extremely difficult, more so than with products: although that’s already very hard. The space of possible artworks is huge, so computational confirmation of novelty (originality and surprise) will be extremely difficult.

My belief is that to be interesting a Computational Artistic Creativity system must be able to either evaluate its own output for creativity (or at the very least for quality), or, more importantly, be able to evaluate its partial solutions for creative potential so that it can choose between possibilities. For Computational Design Creativity systems that design products I think that this is possible: for Computational Artistic Creativity systems I have very grave doubts.

References

- Abrams MH (1953) Orientation of Critical Theories. From: *The Mirror and the Lamp*. In: *Twentieth Century Literary Theory: an Introductory Anthology*. V. Lambropoulos and D. N. Miller (Eds.), 1986, Albany: SUNY Press, pp. 3-31.
- Amabile TM (1983) The Social Psychology of Creativity, *Journal of Personality and Social Psychology*, 43: 997-1013.
- Besemer SP and Treffinger DJ (1981) Analysis of Creative Products: Review and Synthesis, *Journal of Creative Behavior*, 15: 158-178.
- Besemer SP (2006) *Creating Products in the Age of Design*. New Forums Press, Inc.
- Boden MA (1994) What is Creativity?, in MA Boden (ed.), *Dimensions of Creativity*. The MIT Press, pp. 75-117.
- Brown DC (March 2008) Guiding Computational Design Creativity Research, *Proc. NSF International Workshop on Studying Design Creativity'08*, University of

Provence, France. <http://www.cs.wpi.edu/~dcb/Papers/sdc08-paper-Brown-25-Feb.pdf>

Colton S (2008) Creativity versus the Perception of Creativity in Computational Systems. *Proc. AAAI Spring Symposium on Creative Systems*.

Eysenck HJ (1994) The Measurement of Creativity, in MA Boden (ed.), *Dimensions of Creativity*. MIT Press, pp. 199-242.

Horn D and Salvendy G (2006) Consumer-Based Assessment of Product Creativity: A Review and Reappraisal, *Human Factors and Ergonomics in Manufacturing*, 16(2): 155.

Jirousek C (1995) *Art, Design and Visual Thinking*. <http://char.txa.cornell.edu/>, accessed 1st June 2009.

Minsky M (2006) *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster.

O'Quin K and Besemer SP (1989) Development, Reliability, and Validity of the Revised Creative Product Semantic Scale, *Creativity Research Journal*, 2: 267-278.

Russell SJ and Norvig P (2003) *Artificial Intelligence: A Modern Approach*. 2nd edition. Prentice Hall.

Spector L (ed.) (2008) Genetic Programming for Human-Competitive Designs, Special Issue, *AI EDAM*, 22(3).

Thornton S (2008) *Seven Days in the Art World*. W. W. Norton & Co., Inc.

Wikipedia (2009) Art, <http://en.wikipedia.org/wiki/Art>, accessed 1st June 2009.