

A Simple, but Challenging, Model of Computational Design Creativity Evaluation

David C. Brown

Computer Science Department
Worcester Polytechnic Institute
Worcester, MA, USA
dcb@cs.wpi.edu

Abstract

This paper presents a simple model of computational design creativity evaluation, presenting its components with rationale. Components are linked to recent research. The model assumes that the product, not the process, is being evaluated, and that evaluation is done by comparison with descriptions of existing products using a set of aspects that each suggest creativity. Not every evaluation will use all of the components. The model can be used to guide or assess design creativity research.

Introduction

This model is concerned with the Computational Design Creativity (CDC) of engineered products, and specifically creativity evaluation. That is, how does a computer system know when it “sees” a product that people will tend to label it as creative? A key issue is what the appropriate evaluation methods and measures are. In general, the issue is to determine what is involved in such an evaluation: hence the development of the model. The model consists of components, but not every evaluation will use all of the components.

There is no such thing as a “creative computational design system”, only one that produces artifacts that are evaluated as creative. This suggests that any CDC system must design for evaluation. That is why this topic is so important.

Products, design descriptions, and design processes, are labeled as creative based on evaluation [Boden 1994; Hennessey & Amabile 2010]. This model is not concerned with processes, but it is possible that it might apply to them. We assume that evaluation is done by comparison with descriptions of products, using a set of aspects, where each aspect may suggest creativity.

At this point it is still safe to say that humans are better creativity evaluators than machines [Amabile 1996; Hennessey & Amabile 2010], and that (as with much of AI) the best initial approach to computational creativity evaluation of designs is to firmly base it on whatever we can determine about the way that humans do evaluation.

There are many different factors that play a part in evaluation. For example, the time at which the evaluation is done is important for a CDC system. What varies is how much of a design description is available. During designing

it may be partial. After designing it should be complete, but the requirements may not be available, causing it to be much harder to evaluate relative to original intentions.

Evaluation of partial designs, or design decisions, made *during* designing will need to be in terms of their likely contribution to the eventual perceived creativity of the final product. As this is difficult to predict, and evaluation requires accurate expectations [Greco & Brown 2000], partial designs are hard to evaluate.

Of course, even evaluation for creativity *after* the product has been designed is not going to be easy. However, creativity evaluation of sub-parts and sub-systems during designing seems necessary in order to help drive the process towards a creative conclusion. Consequently, evaluations both during and after are needed for CDC systems.

The model proposed here is “simple” in that it provides a framework that has only a few components, but it is “challenging” because to do computationally all that it suggests is currently very difficult. We expect it to remain difficult for a while. However, the model should encourage researchers to try to implement it, and it should allow researchers to classify how evaluation is done in existing and planned CDC systems.

The references provide in this paper are a resource that should allow easy access to the current literature on design creativity evaluation, focusing primarily on the product, hardly at all on the process, and even less on the designer’s personality [Eysenck 1994; Charyton et al. 2008].

We continue by presenting the simple model, and then offer explanations of its components.

The Model

The proposed model of creativity evaluation has the following components:

1. a description of the complete or partial artifact being judged, and/or the actual artifact;
2. the agent judging (person, computer program, or group);
3. the temporal basis for comparison (e.g., the point in time or the period);
4. the source of the basis for comparison (e.g., personal, group, industry, global);

5. the set of “aspects” to include in the evaluation (e.g., novelty, surprise, style, utility, etc.);
6. the method of evaluation for each aspect;
7. the method used to combine the evaluations of the aspects (if one exists);
8. domain knowledge used by the evaluator (i.e., their amount of domain expertise);
9. knowledge about the designer (e.g., performance norms for their level of expertise);
10. knowledge about the audience at whom the evaluation is aimed;
11. knowledge of the design requirements;
12. knowledge of resource constraints (e.g., materials, or design time);
13. the evaluator’s knowledge of the artifact due to the type and duration of experience with it;
14. the evaluator’s knowledge of the design process;
15. the emotional impact of the design on the evaluator;
16. other contextual factors that may have an impact.

An Explanation of the Model

Creativity evaluation depends on the components listed above. We will add some explanation about each one in turn. No detailed consideration will be given here as to how easily each might be adopted, adapted and implemented for CDC system use.

A description of the complete or partial artifact being judged, and/or the actual artifact: The evaluator will judge a design or a partial design. For CDC systems we’re dealing with descriptions, although it is possible that, in the future, CDC systems might be ‘grounded’ by visual and tactile ability that could be applied to (perhaps computer generated) prototype artifacts. Humans are more likely to deal with artifacts, but can also judge descriptions. For complete evaluation it is necessary to have multi-level descriptions (e.g., showing subsystems), and descriptions in terms of Function, Behavior and Structure [Erden et al. 2008]. Some work on creativity evaluation considers a set of designs from a single designer (e.g., in response to the same requirements). However, even though the judgment is about the set, the essence of this approach is still comparing a single design against others.

The agent judging: A human ‘judge’ of a design for creativity might be a person, or group. A CDC system might have knowledge and reasoning based on either.

The temporal basis for comparison: The temporal basis is a point in time, or a period, on which to base the samples of related objects, prototypes, or standards [Redelinghuys 2000] that are used for comparison with the design being judged [Wiggins 2006]. This is especially important for evaluating novelty, for example. The judgment of

creativity is a moving target, as any new artifact could be added to the basis for comparison, which changes any subsequent judgment of the same (or similar) artifact. Of course, that depends on the judging agent having access to the modified basis [Sosa & Gero 2005].

Creativity evaluation is always a judgment at a time. It can be, and usually is, set to “now”, but it could be set in the past, yielding a hypothetical evaluation about whether an artifact might have been seen as creative at some past time. For a CDC system we’re considering “now” to be at the time of designing. By setting both the temporal and the source bases appropriately, evaluations of “rediscoveries” can be made [Sternberg et al. 2002]. The basis might be sourced from a time period. The normal period tends to be the maximal one of all history: at least back to the point where the technology makes comparisons irrelevant (e.g., laser cutters compared to flint knives).

The source of the basis for comparison: This component refers to from where the basis is gathered. It might be personal, in which case the basis is only designs produced by the designer. This corresponds to evaluating for Boden’s P-Creative designs, where P stands for Psychological [Boden 1994]. By widening it to a group, industry, or global, and by using “all history” as the temporal basis, we are evaluating for H-Creative designs, where H stands for Historical. This makes it clear that P- and H-creativity are labels for very particular areas of the time-and-source space of possible bases for comparison: i.e., just referring to P- and H-creative is too simple.

In contrast to the evaluation of a single design against past designs, which might be called “absolute” creativity, some researchers evaluate a design, or a set of designs, against designs produced (often at the same time and from the same requirements) from other designers in the same cohort [Oman et al. 2013; Shah et al. 2003; Kudrowitz & Wallace 2012]. This is often associated with the evaluation aspects of the quantity and variety of ideas generated. This might be called “relative” or “comparative” creativity. However, both can be accounted for using the time and source components in this model.

The set of “aspects” to include in the evaluation: There are a variety of different aspects mentioned in the literature that might be included for creativity evaluation, such as novelty, surprise, style, functionality, and value [Christiaans 1992; Shah et al. 2003; Dean et al. 2006; Horn & Salvendy 2006; Ritchie 2007; Srivathsavai et al. 2010; Liikkanen et al. 2011; Sarkar & Chakrabarti 2011; Kudrowitz & Wallace 2012; Maher & Fisher 2012; Lu & Luh 2012].

Besemer [2006] has one of the most long-lived (from 1981) and complete lists of aspects organized into categories. She includes *Novelty* (Surprising, Original), *Resolution* (Logical, Useful, Valuable, Understandable), and *Style* (Organic, Well-crafted, and Elegant). Cropley &

Kaufman [2012] go even further, proposing 30 indicators of creativity that they experimentally reduced to 24. Their categories of aspects include *Relevance & Effectiveness* (Performance, Appropriateness, Correctness), *Problematization* (Prescription, Prognosis, Diagnosis), *Propulsion* (Redefinition, Reinitiation, Generation, Redirection, Combination), *Elegance* (Pleasingness, Completeness, Sustainability, Gracefulness, Convincingness, Harmoniousness, Safety), and *Genesis* (Vision, Transferability, Seminality, Pathfinding, Germinality, Foundationality).

The method of evaluation for each aspect: Whichever aspects are included in a CDC system, an actual evaluation needs to be made using those aspects [Brown 2013]. For example, an artifact needs to be judged for its novelty/originality [Lopez-Mesa & Vidal 2006; Shelton & Arciszewski 2007; Srivathsavai et al. 2010; Sarkar & Chakrabarti 2011; Maher & Fisher 2012; Brown 2013a] or for whether it is surprising [Brown 2012; Macedo et al. 2009]. Different evaluation methods are possible for both of these aspects. We conjecture that this will be true for other aspects as well. It may be possible to apply the evaluation of aspects to different levels of abstraction in the description [Shah et al. 2003] [Nelson et al. 2009] [Farzaneh et al. 2012], and to descriptions of Function, Behavior and Structure [Sarkar & Chakrabarti 2011].

The method used to combine the evaluations of the aspects: Evaluations have strengths; therefore artifacts may be seen as more, or less, creative – i.e., it isn't a Boolean decision. However, with many aspects being evaluated this will produce a 'profile' of the amount of creativity demonstrated in each aspect. Evaluation in a single, combined dimension results from the evaluator's biases about how to combine different aspect evaluations [Shah et al. 2003; Oman et al. 2013; Sarkar & Chakrabarti 2011; Ritchie 2007]. For a particular agent being modeled, this combination method may not exist. A complex issue that needs addressing is how the separate evaluations of creativity in Function, Behavior and Structure affect each other and the evaluation of the whole artifact.

The domain knowledge used by the evaluator: It is well established in the literature that the amount of domain expertise that the designer has makes a big difference to their potential for creativity. However, to *fully* appreciate a design the evaluator needs to (at least) match their level of sophistication. For example, expert evaluators may know about complex electromechanical devices: less expert designers may only know about Legos. Hence the nature and amount of the evaluator's domain knowledge will make a big difference to the evaluation [Cromptley & Kaufman 2012]. Note that this need not be put explicitly into a CDC system – in fact it may not be able to be – but it might be accumulated using machine learning.

The knowledge about the designer: Knowledge of the capabilities of the designer may play a role in creativity evaluation. Also, knowing the performance norms for the designer's level of expertise is important. Consider a design description of a building from a 10 year old child versus a design description from an excellent Architect. An excellent kid might be very creative relative to what they've already done (P-Creative), while an excellent architect is more likely to be judged as very creative relative to what everyone else has already done (H-Creative). Given knowledge about the designer, there's also the possibility that the evaluator might be able to recognize Transformational creativity [Boden 1994; Ritchie 2006].

The knowledge about the audience at whom the evaluation is aimed: The judgment must be understandable by the recipient of the evaluation. What you'd tell a child would be different from what you'd tell an expert. The conjecture is that this is not just a matter of the type of language used for the evaluation report, but that the actual evaluation might vary. For example, if a simple Yes/No or numeric position on a scale answer is desired then a powerful general technique such as CSPs, Neural Nets, or Evolutionary Computing might be used for the evaluation, as rationale for either the design or the evaluation is not needed, nor available. If the evaluation is for an expert, then it might be provided in technical terms, and mention features, for example: whereas an evaluation of a process might mention ingredients such as selection, planning, evaluation, constraint testing, patching, failure handling, etc.

The knowledge of the design requirements: Do the 'requirements' for the product, possibly including the intended function, need to be known to evaluate creativity? We argue that it is not *necessary*, but it should be helpful, as it allows the basis for comparison to be more precisely selected.

The knowledge of resource constraints: If an evaluator understands how a designer dealt with resources constraints, such as limits on materials availability or limited design time, it can affect their creativity evaluation.

The evaluator's knowledge of the artifact due to the type and duration of experience with it: An evaluator might read the design description, see the artifact, touch the artifact, manipulate the artifact, or actually use the artifact [Ludden et al. 2008]. This affects the completeness of their understanding of the artifact, and therefore their evaluation. Ludden's example involves 'surprise', but other aspects could also be affected.

The evaluator's knowledge of the design process: Colton [2008] argues that, especially for art or artistic products, knowledge of the process is extremely important for the evaluation of creativity. However, it is clear that a very novel and interesting process might result in a not very creative design. We include this component of the model for completeness. For many of the researchers referenced in this paper, this component isn't essential.

The emotional impact of the design on the evaluator: What is the role of emotion in the evaluation of creativity? The "impact" on the evaluator does play a role [Christiaans 1992] but how do fun, cuteness, cleverness, memories, or jokes play role in evaluation? Horn & Salvendy [2006] claim that arousal and pleasure influence the evaluation of product creativity. Cropley [2009] points out that "departure from the usual arouses discomfort" and perhaps "departure from the usual arouses excitement". What isn't clear is how this component might be incorporated into a CDC system.

Other contextual factors that may have an impact: This, we must admit, is a catch-all category. However, there are factors that may play a role in evaluation that could go here, as it isn't clear that they always apply or are a main influence. One such factor is whether a past artifact has been acknowledged as creative: perhaps to the point of it being a disruptive product, changing the direction of future artifacts in the same category. Sternberg et al. [2002] describes this as "propelling" a field. This knowledge might be used to suggest that a new artifact might be creative by analogy: if the new artifact (X) has 'similar' characteristics to an existing artifact (Y), and Y was seen as creative and influential in the past, then perhaps X will be seen as a creative influence. Such an evaluation would be helped by having design time and similarity information [Minsky 2006] available.

Some evaluation schemes include "usefulness" and the "importance" of the use as evaluation aspects (e.g., [Sarkar & Chakrabarti 2011]). This might be measured in terms of actual use, or potential use. As the artifact has just been, or is still being, designed, evaluating "actual" use will not be possible. There needs to be enough knowledge included during the design creativity evaluation process to estimate how much it might be used, and weight it by "importance" or potential "impact".

Summary & Conclusion

This model is intended to be used to guide or assess design creativity research, with the hope that it will eventually apply to CDC systems. The references should allow easy access to the current literature on design creativity evaluation. Given the number and difficulty of the components in the model it is obvious that CDC systems still need a lot of work. The model also makes it clear that

how creative an artifact is may only be properly stated if the full context of the evaluation is included.

References

- T.M. Amabile (1996) *Creativity in Context*, Westview Press.
- S.P. Besemer (2006) *Creating products in the age of design. How to improve your new product ideas!* New Forums Press, Inc.
- M.A. Boden (1994) "What is Creativity?", *Dimensions of Creativity*, M. A. Boden (Ed.), The MIT Press, pp. 75-117.
- D.C. Brown (2012) "Creativity, Surprise & Design: An Introduction and Investigation", *2nd Int. Conf. on Design Creativity (ICDC2012)*, Glasgow, UK.
- D.C. Brown (2013) "Guiding Computational Design Creativity Research", *Int. Jnl. of Design Creativity and Innovation 1(1)*.
- D.C. Brown (2013a) "Variations on Newness", in preparation.
- C. Charyton, R. J. Jagacinski & J.A. Merrill (2008) "CEDA: A Research Instrument for Creative Engineering Design Assessment", *Psychology of Aesthetics, Creativity, and the Arts 2(3)*, 147-154.
- H.H.C.M. Christiaans (1992) *Creativity in Design: The role of domain knowledge in designing*, Uitgeverij Lemma BV.
- S. Colton (2008) "Creativity versus the perception of creativity in computational systems", *Proceedings of the AAAI Spring Symposium on Creative Systems*.
- D.H. Cropley (2009) "Fostering and measuring creativity and innovation: individuals, organisations and products", *Proc. Conf. Can Creativity be Measured?* section 17, Education and Culture DG, European Commission, http://ec.europa.eu/education/lifelong-learning-policy/creativity-conference_en.htm
- D.H. Cropley & J.C. Kaufman (2012) "Measuring Functional Creativity: Non-Expert Raters and the Creative Solution Diagnosis Scale", *Journal of Creative Behavior*, 46(2), 119-137.
- D.L. Dean, J. M. Hender, T.L. Rodgers & E.L. Santanen (2006) "Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation", *Jnl. Association for Information Systems 7(10)*.
- M.S. Erden, H. Komoto, T.J. van Beek, V. D'Amelio, E. Echavarria & T. Tomiyama (2008) "A Review of Function Modeling: Approaches and Applications", Special issue on Multi-modal Design, A. Goel, R. Davis & J.S. Gero (Eds.), *AIEDAM*, 22(2).
- H.J. Eysenck (1994) "The Measurement of Creativity", *Dimensions of Creativity*, M. A. Boden (Ed.), The MIT Press, pp.199-242.

- H.H. Farzaneh, M.K. Kaiser, B. Schroer, V. Srinivasan & U. Lindemann (2012) "Evaluation of Creativity: Structuring Solution Ideas Communicated in Groups Performing Solution Search", *Int. Design Conf.*, DESIGN 2012, pp. 1871-1880.
- D.L. Grecu & D.C. Brown (2000) "Expectation Formation in Multi-Agent Design Systems", *Proc. Artificial Intelligence in Design '00*, J. S. Gero (Ed.), Kluwer.
- B.A. Hennessey & T.A. Amabile (2010) "Creativity", *Ann. Rev. Psychology* 61, 569-598.
- D. Horn & G. Salvendy (2006) "Product creativity: conceptual model, measurement and characteristics", *Theoretical Issues in Ergonomics Science* 7(4), 395-412
- B.M. Kudrowitz & D.R. Wallace (2012) "Assessing the Quality of Ideas from Prolific, Early-Stage Product Ideation", *Jnl. Engineering Design*, Special Issue on Design Creativity, (ifirst online preview), 1-20.
- L.A. Liikkanen, M.M. Hämäläinen, A. Häggman, T. Björklund, & M.P. Koskinen (2011) "Quantitative Evaluation of the Effectiveness of Idea Generation in the Wild", in *Human Centered Design*, Lecture Notes in Computer Science, 6776, pp. 120-129.
- B. Lopez-Mesa & R. Vidal (2006) "Novelty Metrics in Engineering Design Experiments", *Int. Design Conf.*, DESIGN 2006, pp. 557-564.
- C-C. Lu & D-B. Luh (2012) "A Comparison of Assessment Methods and Raters in Product Creativity", *Creativity Research Jnl.*, 24(4), 331-337.
- G.D.S. Ludden, H.N.J. Schifferstein & P. Hekkert (2008) "Surprise As a Design Strategy", *Design Issues* 24(2), 28-38.
- L. Macedo, A. Cardoso, R. Reizenzein, E. Lorini & C. Castelfranchi (2009) "Artificial Surprise", *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*, J. Vallverdu & D. Casacuberta (Eds.), Information Science Reference (IGI Global).
- M.L. Maher & D.H. Fisher (2012) "Using AI to Evaluate Creative Designs", *Proc. 2nd Int. Conf. on Design Creativity (ICDC2012)*. Glasgow, UK, pp. 45-54.
- M. Minsky, (2006) *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, Simon & Schuster.
- B.A. Nelson, J. Yen, J.O. Wilson & D. Rosen (2009) "Refined Metrics for Measuring Ideation Effectiveness", *Design Studies* 30, 737-743.
- S.K. Oman, I.Y. Tumer, K. Wood & C. Seepersad (2013) "A comparison of creativity and innovation metrics and sample validation through in-class design projects", *Research in Engineering Design* 24(1), 65-92.
- C. Redelinghuys (2000) "Proposed criteria for the detection of invention in engineering design", *Jnl. Engineering Design* 11(3), 265-282
- G. Ritchie, (2006) "The transformational creativity hypothesis", *New Generation Computing* 24, 241-266.
- G. Ritchie (2007) "Some Empirical Criteria for Attributing Creativity to a Computer Program", *Minds & Machines* 17, 67-99.
- P. Sarkar & A. Chakrabarti (2011) "Assessing design creativity", *Design Studies* 32, 348-383
- J.J. Shah, N. Vargas Hernandez, S. M. Smith (2003) "Metrics for measuring ideation effectiveness". *Design Studies* 24, 111-134.
- K.A. Shelton, T. Arciszewski (2007) "Formal innovation criteria", *Int. Jnl. Computer Applications in Technology* 30(1/2), 21-32
- R. Sosa & J.S. Gero (2005) "A computational study of creativity in design: the role of society", *AIEDAM* 19(4), 229-244.
- R. Srivathsavai, N. Genco, K. Hölttä-Otto & C.C. Seepersad (2010) "Study of Existing Metrics Used in Measurement of Ideation Effectiveness", *Proc. ASME 2010 Int. IDETC & CIE Confs.*, DETC2010-28802, Montreal, Canada.
- R.J. Sternberg, J. C. Kaufman & J. E. Pretz (2002) *The Creativity Conundrum: A propulsion model of kinds of creative contributions*, Psychology Press.
- G.A. Wiggins (2006) "A Preliminary Framework for Description, Analysis and Comparison of Creative Systems", *Jnl. of Knowledge Based Systems* 19(7), 449-458.