Evaluating Creativity in Humans, Computers, and Collectively Intelligent Systems

Mary Lou Maher
Design Lab
University of Sydney
Sydney, NSW, Australia 2006
marylou.maher@sydney.edu.au

ABSTRACT

Creativity studies focus on the processes that produce creative artifacts and how we evaluate an artifact to determine if it is creative. This paper focuses on the essential criteria in evaluating if a potentially creative artifact is creative. Evaluating creativity is still largely subjective and not well supported with computational tools. An evaluation metric is presented as a way of measuring three essential criteria for creativity: novelty, value, and unexpectedness. The metric is independent of the domain or discipline and does not depend on whether the system producing the creative artifact is a person, a computer, or a combination of human and computer agents. Novelty is a measure of the distance from other artifacts in the space, characterizing the artifact as similar but different. To distinguish this from novelty, value is a measure of the artifact's performance or acceptance rather than a measure of how the artifact's description differs from other artifacts in its class. A metric for value has to accommodate that a creative artifact can change the value system by introducing a performance or function that did not exist in the class of known artifacts. Unexpectedness is measured by how far the artifact is from the expected next artifact.

Keywords

evaluating creativity, novelty, value, unexpectedness, human-computer creativity

INTRODUCTION

Creativity is a topic of philosophical and scientific study considering the scenarios and human characteristics that enable creativity as well as the properties of computational systems that exhibit creative behavior. When studying creativity, we can study how creativity occurs focusing on the *processes* that produce creative artifacts and we can study what makes an act creative focusing on how we *evaluate* an artifact to determine if it is creative. These studies focus on human creativity (eg psychology studies) or computational creativity (eg philosophical studies and artificial intelligence studies). The study of human creativity tends to focus on the characteristics and cognitive behavior of creative people and the environment or situations in which creativity is facilitated. The study of computational creativity, while inspired by concepts of human creativity, is often expressed in the formal language of search spaces and algorithms. The increasing use of computational systems as creative agents and as the basis for crowdsourcing human creativity blurs the boundaries between human and computational creativity. This paper presents a common approach for evaluating the creativity of an artifact across domains and sources, accommodating a way of measuring creativity in a broad range of new human-computational systems.

Why do we need a metric that is independent of the domain of the creative act, or the entity, or process that is being creative? Firstly, there is an increasing interest in understanding computational systems that can formalize or model creative processes and therefore exhibit creative behaviors or acts, yet our best example of creative entities are human. In parallel there is increasing interest in computational systems that encourage and enhance human creativity that make no claims about whether the computer is being or could be creative. Finally, as we develop more capable socially intelligent computational systems and systems that enable collective intelligence among humans and computers, the boundary between human creativity and computer creativity blurs. As the boundary blurs, we need to develop ways of evaluating or recognizing creativity that makes no assumptions about whether the creative entity is a person, a computer, a potentially large group of people, or the collective intelligence of human and computational entities. This paper presents a metric that can be adapted and applied to the various situations in which creativity is being studied, providing a basis for comparison across domains and computational processes.

CREATIVE PROCESSES AND EVALUATION

This paper makes a distinction between studying and describing the processes that generate potentially creative artifacts, which focus on the cognitive behavior of a creative person or the properties of a computational system, and the essential criteria for evaluating if a potentially creative artifact is creative. In this paper *creative artifact* is

DESIRE '10, 16-17 August 2010, Aarhus, Denmark

a term that refers to the result of creativity in any field, whether artistic, design, mathematical, or science. In the description of the evaluation metric, an assumption is made that an artifact can be described as a set of attribute-value pairs. Artifacts may be have structured descriptions as attribute-value pairs, but may also be described as images, unstructured text, 3D models, etc. The use of attribute-value pairs as the basis for evaluation is exemplary, but not limiting. There are many fields in which the creative artifact cannot be described as attribute-value pairs or decomposed into discrete parts. The metric described below can be reformulated and adapted for other ways of representing or describing artifacts.

When describing creative processes there is an assumption that there is a space of possibilities. Boden [3] refers to this as conceptual spaces and describes these spaces as structured styles of thought. In computational systems such a space is called a state space. How such spaces are changed, or the relationship between the set of known artifacts, the space of possibilities, and the potentially creative artifact, is the basis for describing processes that can generate potentially creative artifacts.

There are many accounts of the processes by which a potentially creative artifact can be produced. Two sources described here are: Boden [3] from the philosophical and artificial intelligence perspective and Gero [6] from the design science perspective. The processes for generating potentially creative artifacts are described generally by Boden [3] as three ways in which creative artifacts can be produced: combination, exploration, and transformation: each one described in terms of the way in which the conceptual space of known artifacts provides a basis for producing a creative artifact and how the conceptual space changes as a result of the creative artifact. Computational processes for generating potentially creative designs are articulated by Gero [6] as combination, transformation, analogy, emergence, and first principles. These processes can become operators for generating artifacts that explore, expand or transform the relevant state space. While these processes provide insight into the nature of creativity and provide a basis for computational creativity, they have little to say about how we know if the result of the process, a potentially creative artifact, is in fact creative. As we move towards computational systems that enhance or contribute to human creativity, the articulation of process models for generating creative artifacts does not provide an evaluation of the product of the process and are insufficient for evaluating if a potentially creative artifact is creative. Such systems that generate potentially creative artifacts need a model of evaluation that is independent of the process by which the artifact was created.

A common claim for computational creativity is based on the distinction between P-creativity (psychological) and H-creativity (historical) [3], where computers can be P-creative. P-creativity is a creative artifact that is novel for the individual or computer that produced it and H-creativity is novel historically. When we consider the evaluation of potentially creative artifacts that are generated by humans, computers, or combinations of humans and computers, it will be increasingly difficult to determine the boundary of the state space that is the basis for P-creativity. The evaluation model in this paper assumes there is a relevant state space of artifacts associated with the potentially creative artifact. This state space is not bounded before the process for producing the potentially creative artifact begins and can include an initially fixed state space representation, personal knowledge, historical knowledge, or the knowledge available to a network of humans and computers. In this paper, the evaluation metrics are independent of the distinction between P-creativity and H-creativity.

ESSENTIAL CRITERIA FOR EVALUATING CREATIVITY

The evaluation metric presented in this paper is based on three essential criteria for evaluating if an artifact is creative: The artifact is novel, valuable, and unexpected. These three criteria are introduced below and compared to other approaches to recognizing or evaluating potentially creative artifacts.

Novel: Novelty is a measure of how different the artifact is from known artifacts in its class. Generally, artifacts are put in a class according to their label or function, eg a chair or a car. Members of a class are similar across their attributes and vary according to the values of the attributes. The attributes may be further classified, for example, as structure, behavior, function [6], but this does not change the evaluation of novelty. Novelty is recognized when a new attribute is encountered in a potentially creative artifact, a previously unknown value for an attribute is added, or a sufficiently different combination of attributes is encountered. Novelty can be measured as a distance from other artifacts in the space, characterizing the artifact as similar but different.

Value: Value is a measure of how the potentially creative artifact compares to other artifacts in its class in utility, performance, or attractiveness. Often this is a measure of how the artifact is valued by the domain experts for this class of artifact and is either a weighted sum of performance attributes or is a reflection of the acceptance of this artifact by society. To distinguish this from novelty, value is a measure of the artifact’s performance rather than a measure of how the artifact’s description differs from other artifacts in its class. When an artifact is described by a set of attributes, it is possible that some of the attributes are performance attributes, and so some of the information for measuring value may be embedded in the description. Defining a fixed metric for value is not possible because often a creative artifact can change the value system by introducing a performance or function that did not exist in the class of known artifacts before the creative artifact. Therefore, the metric for value needs to be an adaptive function.

Unexpected: The measurement for unexpectedness has to do with the recent past and how we develop expectations for the next new artifact in a class. This is distinguished from novelty because it is based on tracking the progression of one or more features in a class of artifacts, and changing the expected next difference. The amount of difference is not relevant as it is in the novelty metric, the variation from expectation is relevant.

Most definitions or evaluation of creativity, including definitions in the dictionary, will include novelty as an essential part of the definition. Some definitions will state that value is the umbrella criteria and novelty, quality, surprise, typicality, and others are ways in which we characterize value for creative artifacts. For example, Boden [3] claims that novelty and value are the essential criteria and that other aspects, such as surprise, are kinds of novelty or value. Wiggins [21] often uses value to indicate all valuable aspects of a creative artifact, yet provides definitions for novelty and value as different features that are relevant to creativity. Oman and Tumer [15] combine novelty and quality to evaluate individual ideas in engineering design as a relative measure of creativity. Shah, Smith, and Vargas-Hernandez [15] associate creative design with ideation and develop metrics for novelty, variety, quality, and quantity of ideas. In this paper, novelty and value are presented as distinct features of a creative artifact: novelty is based on a comparison of a description of the potentially creative artifact to other artifacts and value is a derivative feature that requires an interpretation of the potentially creative artifact from outside the description of the artifact.

Several researchers consider unexpectedness, or surprise, to be a relevant feature of creativity. Wiggins [21] argues that surprise is a property of the receiver of a creative artifact, that is, it is an emotional response. Wiggins' view of surprise is similar to the definition of value because the interpretation lies outside the description of the artifact. Boden [3] claims that surprise is a kind of novelty. In this paper, surprise is a separate essential criterion for evaluating a potentially creative artifact because it is possible for something to be novel and valuable, but not be surprising. Surprise is a feature that is based on expectations and so is a function of the attributes of the potentially creative artifact in comparison to other artifacts (like novelty), but also depends on a projection or expected value that lies outside the description of the artifacts (like value). Since unexpectedness is associated with creativity and is different operationally from both novelty and value, then novelty and value are not sufficient.

Ritchie [16] has two essential criteria for creativity: novelty and quality. These roughly correspond to the definitions of novelty and value in this paper. Ritchie elaborates on novelty to include typicality as an essential feature, and further claims that such primitive elements can only be judged by people. In this paper, the three criteria are further formalized as a starting point for developing a common metric for creativity across domains and so creativity can be "judged" by humans, computers, and/or human-computer systems.

EVALUATION METRIC FOR CREATIVITY

Often a formalization of creativity starts with a space of possibilities and the properties of a person or computational system that can produce an artifact within that space that is creative. Many assume that the evaluation of the artifact as creative is determined by people (individual judges, gatekeepers, society), or is assumed when the system that produced the artifact has the properties of a creative system. In this section, the definitions of novelty, value, and unexpected are further specified as a metric for evaluating the potentially creative artifact with some examples of computational approaches to evaluating creativity.

If the space of possibilities is a universal space, U, then there is a subset of that space, C, which describes a class of artifacts that characterizes the known artifacts in that class. A subset of the class of artifacts, A, includes the known set of artifacts.

$$A = \{a_1, a_2, \ldots, a_n\} . \qquad (\mathbf{1})$$

For the purposes of describing the evaluation metric, a_i is a new and potentially creative artifact. The evaluation metric, E, is a function of a_i.

$$E(a_i) = N(a_i) \times V(a_i) \times S(a_i) . \qquad (\mathbf{2})$$

where
a_i is creative if $E(a_i) > 0$
N, V, and S are functions that return a value >= 0
N is a measure of the novelty of a_i
V is a measure of the value of a_i
S is a measure of the unexpectedness, or amount of surprise of a_i
The relative importance of each of the three criteria is not presumed in this formulation of the evaluation function since that would be dependent on the source of the creative artifact and the field to which it contributes (art, science, etc.). The measure of each of the criteria is as important to evaluating a potentially creative artifact as the final numerical value of the metric, E.

Evaluating novelty: N

$$N(a_i) = f(d(a_i)). \qquad (\mathbf{3})$$

Novelty is a measure of the distance, d, between the potentially creative artifact and the other artifacts in the class. Calculating the distance from each known artifact is a start, but doesn't provide a way of characterizing the distance relevant to a set of artifacts that may be scattered around a potentially large state space. The important consideration in measuring novelty is finding a measure that can characterize how a new artifact is different from other artifacts in the space. Since artifacts can be described by a potentially large number of attributes and can be distributed throughout a state space, this paper presents clustering approaches to measuring novelty. Various clustering algorithms provide a way of characterizing distance from a group or cluster of artifacts. Two described here are K-means clustering (the algorithm was first published by Lloyd [9]) and Self-Organizing Maps (SOM) [8].

K-means clustering uses a set of centroids to represent clusters of input data, or in our case, clusters of artifacts. In order to use k-means clustering to evaluate the novelty of a potentially creative artifact, k-means clustering partitions n artifacts, $\{a_1, a_2, \ldots, a_n\}$, where each artifact is a d-dimensional vector of attribute-value pairs, into K sets, where k<n and S={S1, S2, …, Sk} such that the within-cluster sum of squares in minimized:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{a_j \in S_i} \left\| a_j - \mu_i \right\|^2 \quad (4)$$

When K-means clustering is used to determine the novelty of a potentially creative artifact, the update function is used to determine how far the new artifact is from the centroid of the most similar cluster. The most similar cluster is selected as the centroid $K_{(t)}$ with the minimum distance d to the potentially creative artifact where d is calculated using the K-means distance function:

$$d(a_i) = \sqrt{\sum \sum_{i=1}^{d} \Delta(k_i, a_i)^2} \quad (5)$$

Alternatively, self-organizing maps (SOMs) provide a way to take an n-dimensional space and map it onto a 2-dimensional space. This simplifies the measurement of distance between 2 points in the space. SOMs comprise a number of neurons that represent clusters of input data, in our case clusters of artifacts in class C. When used to determine novelty, SOM neurons represent the current set of artifacts, A, in class C. The initial condition is a single neuron, and the update function adds a new neuron to the map. The SOM update function progressively modifies each neuron K to model a cluster of artifacts that are relevant to the most recently added artifact, but also influenced by past observations or events. When a potentially creative artifact is presented to the SOM, each neuron is updated by adding randomly initialized variables k_L with any attributes that occur in a_i but not in K. The most similar artifact model is then further updated by selecting the neuron $K_{(t)}$ with the minimum distance d to the input stimulus where d is calculated using the SOM distance function:

$$d(a_i) = \sqrt{\sum_{i} \Delta(k_i, a_{i_i})^2} \quad (6)$$

Similar to the d calculated in the update function for k-means clustering, the d calculated using the SOM distance function is the basis for determining if the potentially creative artifact is creative.

There are many accounts of measuring novelty using computational approaches. Marsland et al. [12] used Stanley's model of habituation [20] to implement a real-time novelty detector for mobile robots. Like the Kohonen Novelty Filter [8], the real-time novelty detector uses a Self-Organising Map (SOM) as the basis for the detection of novelty. Habituation and recovery extends a novelty filter with the ability to forget. This allows novel artifacts that have been seen in the past to be considered again as potentially creative using a new value system.

Saunders and Gero [17] drew on the work of Berlyne [2] and Marsland et al [12] to develop computational models of curiosity and interest based on novelty. They used a real-time novelty detector to implement novelty. However, they were also looking for a way to measure interest, where novelty is not the only determinant of interest. Rather, interest in a situation is also related to how well an agent can learn the information gained from novel experiences. Consequently, the most interesting experiences are often those that are 'similar-yet-different' to previously encountered experiences. Saunders and Gero [17] model interest using sigmoid functions to represent positive reward for the discovery of novel stimuli and negative reward for the discovery of highly novel stimuli. The resulting computational models of novelty and interest are used in a range of applications including curious agents. The use of a sigmoid function to provide negative reward for highly novel artifacts may be relevant as a computational model for novelty that can recognize when an artifact is too different from the known artifacts in the class to be considered creative.

In summary, the measure of novelty for a potential creative artifact is essentially a distance measure. The specific definition of distance can be based on a one of many possible novelty algorithms. The value of novelty is a function of the distance metric, but may require establishing a threshold above which an artifact is considered creative, or a sigmoid function for determining a range for the distance metric as in the Saunders and Gero model [17].

Evaluating Unexpectedness/Surprise: S

$S(a_i(t)) = 1$ if $a_i(t) <> a_n(t)$ in the sequence $(a_1(t\text{-}n), a_2(t\text{-}n+1), \ldots, a_n(t))$; otherwise 0. (**7**)

An artifact, a_i, is considered surprising when we recognize a pattern in recent artifacts, and the potentially creative artifact does not follow the expected next artifact in the pattern.

A class of artifacts establishes expectations for new artifacts in that class. For example, when we think of cars as a class of artifacts, we have expectations about the purpose and value of the car, and many of the structural components of the car. A car design that meets our expectations but also satisfies the novelty criteria may not be considered creative. A creative design for a car takes some aspect that we have come to expect even in novel car designs and changes it. When hybrid cars were first introduced, the car changed our expectations in two ways: while we expected novel electric cars to produce more efficient batteries, the hybrid car uses both gas and stored electric energy and the energy is stored while the car is using gas; while we expected that a novel car design would allow us to drive farther with the same amount of gas, the hybrid car showed that status as an environmentally friendly driver was an equally important value.

A major difference between evaluating novelty and expectation is the sequential nature of expectation. Surprise is achieved by setting up expectations over a period of time or over a sequence of designs. Novelty can be measured without considering the sequence in which the artifacts are generated or experienced. While surprise may be considered a kind of novelty, measuring surprise is distinct in requiring that expectations be established in a sequence of events or acts, and when those expectations are not met we are surprised. Two examples of the sequential nature of expectations and surprise are humor and music.

Recognizing and therefore evaluating surprise requires the identification of patterns, and those patterns can be considered abstractly and separate from the content. In humor, Clarke [5] explains: "The theory is an evolutionary and cognitive explanation of how and why any individual finds anything funny. Effectively it explains that humour occurs when the brain recognizes a pattern that surprises it, and that recognition of this sort is rewarded with the experience of the humorous response, an element of which is broadcast as laughter.

"By removing stipulations of content we have been forced to study the structures underlying any instance of humour, and it has become clear that it is not the content of the stimulus but the patterns underlying it that provide the potential for sources of humour. For patterns to exist it is necessary to have some form of content, but once that content exists, it is the level of the pattern at which humour operates and for which it delivers its rewards."

In music, there is a similar phenomenon where the notes in a musical score set up expectations, and a note that does not meet our expectation is perceived as surprising. Measuring surprise can be achieved with pattern matching and analogical reasoning, by looking for analogous series of artifacts and measuring the distance between the expected next artifact and the potentially creative artifact. This approach is a basis for understanding creativity in music implemented in Musicat [14]. Musicat looks for a series of notes in a musical score that form a group and that are repeated in similar patterns. Once a series of these groups is found, a next group can be compared for similarity. The role of previous groups is to create tension, as surprise is achieved when the elements of the next group do not match the expected sequence.

There is a similar phenomenon in machine learning, where Schmidhuber [18] defines a number of computational models of interest for use in reinforcement learning (RL) systems. One such model uses the predictability of a learned world model to represent curiosity and boredom as reinforcement and pain units. Predictability is measured as the result of classifying sequential observations using a self-supervised neural network. The resulting ‘curious neural controller’ works in conjunction with RL and is designed to reward situations where the model network’s prediction performance is not optimal, in order to encourage an agent to revisit those situations and improve its network model. As in Berlyne’s [2] theory, maximum motivational reward is generated for moderate levels of predictability to represent curiosity about situations in which an ‘ideal mismatch’ occurs between what is expected and what is observed. This reward process is designed to represent the theory that a system cannot learn something that it does not already almost know, very similar to recognizing a potentially creative artifact.

In summary, a metric for unexpectedness is based on a pattern matching algorithm that finds sequences of artifacts that are similar except for one or a few attributes that change in a predictable sequence. If the potentially creative artifact shares the unchanging attributes and does not have the expected value for the attribute that changes, then it is unexpected.

Evaluating Value: *V*

$$v(a_i) = \sum_{j=1}^{n} w_j p_j (a_{i,j}) \quad \textbf{(8)}$$

where

w_i is the weight of performance variable i

$p_j(a_{i,j})$ is the value of performance variable j for a_i

Value is the third essential characteristic of a creative artifact. Novelty and surprise may be present in a potentially creative artifact, but if the artifact has is not valued then it will not be considered creative. The value of a potentially creative artifact can be a social phenomenon and determined by the “gatekeepers” as described by Csikszentmihalyi [4], or it can be codified in an adaptive evaluation function. The value of an artifact is often judged by criteria that are established by the requirements and performance attributes associated with the class of artifacts. This is expressed as a weighted sum of the performance variables of the artifact. The difficulty with this approach is that often a creative artifact will change our value system and introduce new performance variables, and the function associated with value should allow the performance variables to change in response to the potentially creative artifact.

To demonstrate how the measurement of value in evaluating potentially creative artifacts can be an adaptive function, a co-evolutionary genetic algorithm is presented. The genetic algorithm approach introduced by Holland [7] provides a mechanism for searching a state space that evaluates a new population of potential solutions with a fitness function. In a genetic algorithm a new artifact evolves through an iterative process of combining and mutating the genotypes of entities in a state space of possible artifacts. This iterative process continues until an artifact has been generated that satisfies a fitness function. This fitness function is essentially a measure of the value of each newly generated artifact. The basic algorithm is shown below, where A(t) is the space of possible artifacts, ρ is a function that calculates the fitness (or value) of the artifacts in the state space for the current generation of designs, and select, crossover, and mutation are the genetic operators. The termination condition is achieved when an artifact in the current generation reaches a threshold for fitness or a number of iterations.

```
t = 0;
   initialize genotypes in P(t);
   evaluate phenotypes in P(t) for
fitness;
   while termination condition not
satisfied do
      t = t + 1;
      select P(t) from P(t-1);
      crossover genotypes in P(t);
      mutation of genotypes in P(t);
      evaluate phenotypes in P(t);
```

In this algorithm, the fitness function is the measure of the value, $v(a_i)$ of a potentially creative artifact. When the fitness function is allowed to adapt to changing value systems in response to new artifacts, rather than serve as predefined criteria for success, we are able to consider the genetic algorithm as a model for generating and evaluating creative artifacts.

One way to allow the fitness function to adapt to the new generation of artifacts is to have a person serve as the fitness function or allow a person to modify the fitness function in response to artifacts generated by crossover and mutation. Bentley [1] contains many examples of computer generated designs using evolutionary processes as an approach to exploring a search space. McCormack [13] shows how evolutionary algorithms can be guided by a person to generate creative artifacts.

A second way to allow the fitness function to adapt to the new generation of artifacts is to consider the fitness function as a search space. This provides a formal method for changing the fitness function in response to the current generation of artifacts, similar to our experience with people where they may change what they are looking for based on what they find. This has been modeled as a co-evolutionary approach to design, introduced by Maher [10] and developed further for engineering design problems in [11]. Using a state space representation for the space of possible artifacts, A, and a state space representation for the space of possible values, V, the co-evolution of artifacts and values can be expressed as:

$$A_{final} \in \{A_{\rho 1}, A_{\rho 2}, A_{\rho 3}, \ldots A_{\rho n}\},\ A_{\rho i} = \text{best}_{\rho i}\{A_i\} \quad \text{where } \rho_i = f(V_i) \tag{9}$$

$$V_{final} \in \{V_{\rho' 1}, V_{\rho' 2}, V_{\rho' 3}, \ldots V_{\rho' n}\},\ V_{\rho' i} = \text{best}_{\rho' i}\{V_i\} \text{ where } \rho'_i = f(A_i)$$

where,

ρi is a fitness function for the artifacts at time i,

$\rho' i$ is a fitness function for the values at time i,

V is a space of possible values,

A is a space of possible artifacts,

$A\rho i$ is the set of selected artifacts corresponding to the space of Vi as the current focus for evaluating artifacts ρi., $A\rho i \in A$

$V\rho' i$ is the set of selected values corresponding to the space of Ai as the current focus for evaluating values $\rho' i$, $V\rho' i \in V$

Each new generation of the artifact space, A, and value space, V can be generated using the genetic operators, or other iterative processes for searching a space of possibilities. In this way, the measurement of the set of values of any given set of artifacts is adapted in response to the attributes of the current population of artifacts.

In summary, the measurement of the value of a potentially creative artifact is based partly on the requirements and performance variables that are defined before the potentially creative artifact is produced and is adapted in response to the potentially creative artifact. One way to achieve this in a computational system is to use a co-evolutionary algorithm in which the fitness function changes in response to the current population of potential solutions.

CONCLUSIONS

One impediment to the development of metrics for evaluating creativity is the differing expectations in different domains. As pointed out by Ritchie [16], in the art world, painting a picture or writing a poem is often considered creative, even if it is performed in an ordinary manner; in contrast, in the world of science, math, and engineering, creativity is considered to be rare and only occurs when something exceptional has been produced. This may be why a popular definition of creativity associates creativity with the arts, aka the creative arts. A set of essential criteria for evaluating creativity can apply equally well to artistic and scientific creativity, possibly by raising the bar for what is considered creative in the arts, and by clarifying what we mean by creativity in the sciences. Formalizing the essential criteria for evaluating creativity allows us to compare the many different approaches to developing computational systems that enhance creativity and computational systems that are themselves creative. Without a common metric, we can't compare human, computer, and collectively intelligent systems.

This paper presents three essential criteria for evaluating creativity, regardless of the domain or source of creativity:

novelty, unexpectedness, and value. Novelty is typically associated with creativity and is not hard to argue as an essential characteristic of a creative artifact. Most agree that novelty is not a sufficient condition for creativity and therefore adjectives are applied to clarify what kind of novelty is associated with creativity. This paper formalizes novelty as a measure of distance from known artifacts. Unexpectedness is an aspect of creativity that we recognize when we say that something is creative because it surprises us, does not meet our expectations for the next novel artifact in its class. Unexpectedness is measured using pattern matching algorithms that look for variations across one or more attributes in a sequence of designs. Value is a characteristic of creativity that reflects our individual or social recognition that a highly novel, random act or result is not sufficient for us to judge something as being creative. The creative artifact must somehow extend our understanding in a specific field, change our value system, or enhance our lives in some way. Measuring value is based on a set of performance criteria that can be adapted by the introduction of new performance possibilities in a creative artifact.

The contribution of this paper is a common metric for evaluating creativity that is based on three essential criteria for creativity. The paper shows how the common metric is derived from the various definitions and metrics developed in different domains and sources of creativity. The elements of the metric are not new, but the combination of these three essential criteria is presented as necessary and sufficient criteria for creativity. The metric is not intended to be a fixed algorithm for measuring creativity: it is a formalism that can be adapted and applied to generate different algorithms or to guide human judgment of creativity. The value of this formalism lies in how the criteria have been derived and developed from the literature on creativity. Validation cannot be achieved by applying the criteria to a single creative design scenario. Rather, the validation lies in whether the formalism is applied and used to evaluate creativity across different domains and sources of creativity.

REFERENCES

1. Bentley, P.J.: (Contributing Editor) Evolutionary Design by Computers. Morgan Kaufmann Publishers Inc., San Francisco, CA (1999)
2. Berlyne, D.E.: Novelty, Complexity and Hedonic Value. Perception and Psychophysics 8, 279-286 (1970)
3. Boden, M.: The Creative Mind: Myths and Mechanisms, Routledge; 2 edition (2003)
4. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery and Invention, HarperCollins Publishers (1996)
5. Clarke, A.: The Pattern Recognition Theory of Humour, Pyrrhic House (2008)
6. Gero, J.S.: Computational Models of Innovative and Creative Design Processes, Technological Forecasting and Social Change 64, 183-196 (2000)
7. Holland, J.H.: Concerning Efficient Adaptative Systems, In M.C Yovits, G.T. Jacobi, & G.D Goldstein (editors), Self-organizing Systems, SpartanBooks, 215-230 (1962)
8. Kohonen, T.: Self-Organisation and Associative Memory, Springer, Berlin (1993)
9. Lloyd, S.P.: Least Squares Quantization in PCM, IEEE Transactions on Information Theory 28 (2): 129–137 (1982)
10. Author: Creative Design Using a Genetic Algorithm, Computing in Civil Engineering, ASCE, pp 2014-2021 (1994)
11. Author: Modelling Design Exploration as Co-evolution, Microcomputers in Civil Engineering, 11(3), 195-210 (1996)
12. Marsland, S., Nehmzow, U., and Shapiro, J.: A Real-Time Novelty Detector for a Mobile Robot. EUREL European Advanced Robotics Systems Masterclass and Conference (2000)
13. McCormack, J.: Impossible Nature: The Art of Jon McCormack, Australian Centre for the Moving Image (2004)
14. Nichols, E., Hofstadter, D.R.: Musicat: A Computational Model of Creativity in a Musical Domain, NSF CreativeIT Workshop (2009)
15. Oman, S and Tumer, I. The Potential of Creativity Metrics for Mechanical Engineering Concept Design in Proceedings of the 17th International Conference on Engineering Design (ICED'09), Vol. 2, eds: Norell Bergendahl, M.; Grimheden, M.; Leifer, L.; Skogstad, P.; Lindemann, U., pp 145-156. (2009).
16. Ritchie, G.: Some Empirical Criteria for Attributing Creativity to a Computer Program. Minds and Machines, 17(1):67-99 (2007)
17. Saunders, R., Gero, J.S.: Designing For Interest and Novelty: Motivating Design Agents, CAAD Futures 2001, Kluwer, Dordrecht, 725-738 (2001)
18. Schmidhuber, J.: A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers. in: J.A. Meyer, and S.W. Wilson, (Eds.), The International Conference on Simulation of Adaptive Behaviour: From Animals to Animats, MIT Press/Bradford Books, 222-227 (1991)
19. Shah J., Smith S., Vargas-Hernandez N. Metrics for measuring ideation effectiveness, Design Studies, V24(2), 111-134 (2003).
20. Stanley, J.C.: Computer Simulation of a Model of Habituation. Nature 261, 146-148 (1976)
21. Wiggins, G.: A Preliminary Framework for Description, Analysis and Comparison of Creative Systems, Knowledge-Based Systems 19, 449-458 (2006)