# Using AI to Evaluate Creative Designs

**Mary Lou Maher**
HCI Lab
School of Information Studies
University of Maryland
College Park, MD
ml.maher@umd.edu

**Douglas H. Fisher**
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN
douglas.h.fisher@vanderbilt.edu

**ABSTRACT**

A design for a building or product may be considered creative by some person, group, or the general public regardless of what others might think. We survey the literature on characteristics of creative products, particularly of design. We argue that the three characteristics of novelty, value, and surprise are essential for evaluating creativity in design, although these may be augmented with other considerations that are domain dependent or based on individual interpretation. We propose that measures of distance and Bayesian probability can serve in measuring the three features of creativity, and that strategies like clustering can serve to create and organize the conceptual space against which these features are evaluated. There are at least two motivations for formalizing the assessment of creativity: to give an artificial agent an ability to judge creativity, and to give human analysts a uniform means of evaluating creativity in designs, whether the design stems from a single human, a single artificial agent, or a community of agents, human and/or artificial. We illustrate concepts using an example of sustainable design, the Bloom laptop.

**Keywords**

evaluating creativity, novelty, value, surprise, sustainable design

## INTRODUCTION

Creativity is situated and contextualized: we experience a work of art in a specific museum, we learn about a creative proof in a mathematics course, we buy a creative product in an electronics store. As an area of research, studying creative phenomena is a way of finding common patterns across many examples and disciplines of creativity. Another approach to creativity research is to start with generalized models of creativity and find examples that show how the generalizations apply in specific situations.



In this paper, we develop AI models for three characteristics of the products of creativity as a way of understanding how we recognize creativity and as a starting point for evaluating creative designs.

Models of creativity can focus on either the *processes* that produce creative artifacts or how we *evaluate* an artifact to determine if it is creative from either the perspective of human creativity (for example in psychology studies) or computational creativity (for example in 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.

Why do we need a common model for evaluating creativity that is independent of the domain of the creative design or process that is being creative? Firstly, there is an increasing interest in developing computational systems that can model creative processes and therefore generate creative designs, yet our best example of creative entities are human and our only evaluators are humans. 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. Thus, as part of our arsenal for assessing creativity, we want uniform means of comparing designs, be they the products of a single human with or without computing tools, a single artificial agent, or a community of agents, human and/or artificial.

A related but distinct motivation for our formalizations are to take initial steps at imbuing artificial agents with an ability to assess creativity for purposes of evaluating their own designs, but also so that they can be effective collaborators with humans in increasingly sophisticated socially intelligent computational systems.

Generally, we believe that as the boundary between human creativity and computer creativity blurs, 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 removes any bias associated with individual human creativity.

Informed by a survey of literature on assessing creativity, this paper argues that creativity can be evaluated in terms of novelty, value and surprise, which can be adapted and applied to the various disciplines and situations in which creativity is being studied. Our intent is that this will facilitate comparison and progress across domains and computational processes. We illustrate and demonstrate the concepts using the Bloom laptop (Figure 1), which was designed by mechanical engineering students at Stanford University and Aalto University [5]. The laptop was designed for ease of recycling with design requirements such as: minimum number of parts and types of material, modular construction and disassembly, ease of disassembly, minimum disassembly time.

Figure 1. Bloom Laptop Modular Design (Bohbe et al 2010)

**DESCRIBING CREATIVE PROCESSES AND EVALUATING CREATIVITY**

There is a distinction between studying and describing the processes that generate potentially creative designs, which focus on the cognitive behavior of a creative person or the properties of a computational system, and the methods for evaluating a potentially creative design.

A creative design does not arise from a vacuum, and furthermore, it is typically evaluated within the context of a need or desire that is not fulfilled by existing designs in the same class. When researchers describe creative processes there is an assumption that there is a space of possibilities. Boden [6] calls such a space a “conceptual space” and describes these spaces as structured styles of thought. In computational systems such a space is called a state space, in computational models of design such a space is called a design space. How these 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 product can be produced. Two sources described here are: Boden [6] from the philosophical and artificial intelligence perspective and Gero [9] from the design science perspective. The processes for generating potentially creative producta are described generally by Boden [6] as: combination, exploration, and transformation where each one is described in terms of the way in which the conceptual space of known designs provides a basis for producing a creative design and how the conceptual space changes as a result of the creative design.

Computational processes for generating potentially creative designs are articulated by Gero [9] 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. Maher [15] characterizes different computational processes in terms of transformation and exploration and describes a zone of creativity in order to evaluate their potential for generating creative designs.

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 recognize or evaluate creativity in the resulting product of the process. 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. Systems that generate potentially creative artifacts require a model of evaluation that is independent of the process by which the artifact was generated.

A common claim for computational creativity is based on the distinction between P-creativity (psychological) and H-creativity (historical) [6], 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 models are independent of the distinction between P-creativity and H-creativity.

Csikszentmihalyi and Wolfe [8] define creativity as a an idea or product that is original, valued, and implemented. Most definitions of creativity, including definitions in the dictionary, will include novelty as an essential part of the definition. A definition of creativity may focus on novelty as the primary criterion and claim that novelty is expressed as a new description, new value, or a surprising feature of a creative product. Alternatively, many 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. Villalba [25] provides an overview of creativity research and its measurements. Runco [20] presents several authors that define creativity as involving the creation of something new and useful [3, 4, 23, 17, 2]. Boden [6] claims that novelty and value are the essential criteria and that other aspects, such as surprise, are kinds of novelty or value. Wiggins [26] often uses value to indicate all valuable aspects of a creative product, yet provides definitions for novelty and value as different features that are relevant to creativity. Oman and Tumer [18] combine novelty and quality to evaluate individual ideas in engineering design as a relative measure of creativity. Shah, Smith, and Vargas-Hernandez [22] associate creative design with ideation and develop metrics for novelty, variety, quality, and quantity of ideas.

Amabile [1] says it most clearly when she summarizes the social psychology literature on the assessment of creativity: While most definitions of creativity refer to novelty, appropriateness, and surprise, current creativity tests or assessment techniques are not closely linked to these criteria. She further argues that "There is no clear, explicit statement of the criteria that conceptually underlie the assessment procedures." In response to an inability to establish and define criteria for evaluating creativity that is acceptable to all domains, Amabile [1] introduces a Consensual Assessment Technique (CAT) in which creativity is assessed by a group of judges that are knowledgeable of the field. Within this technique, Amabile defines a cluster of features associated with creativity for the judges to rate that are specific to the artistic or verbal artifact being assessed (for example, in an artwork: creativity, novel idea, variations in shapes, complexity, detail). The CAT does not assist in developing a common set of metrics for evaluating creativity but instead provides a common technique for people to judge creativity.

## NOVELTY, VALUE, SURPRISE AS CHARACTERISTICS OF CREATIVE PRODUCTS

Creativity in a space of possible and existing designs is a relative measure. For something to be creative, it is compared to other artifacts in a class of products or processes. The characteristics of creativity that we describe here are defined as a comparison between a potentially creative design and other designs. While others have grouped novelty and value as a single characteristics of creativity, we define novelty and value as two different characteristics of a creative artifact: novelty is based on a comparison of a description of the potentially creative design to other designs and value is a derivative feature that requires an interpretation of the description of the potentially creative design. That is, novelty considers the descriptive attributes and value considers the performance attributes. Surprise is a third characteristic of a creative design 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 based on recognizing patterns or sequences in the space of designs. Surprise 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).

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, for example the Bloom laptop belongs to the class of laptop designs. Members of a class are similar across their attributes and vary according to the values of the attributes. Novelty is recognized when a new attribute is encountered in a potentially creative design, a previously unknown value for an attribute is added, or a sufficiently different combination of attributes is encountered. For example, the Bloom laptop introduced a new way to describe the body of the laptop. Where the Mac laptops have a unibody, the Bloom laptop has a body that is made of easily separable parts. A model for measuring novelty can be based on the distance of the potentially creative artifact from other artifacts in the same conceptual space, measuring how the design is similar but different.

Value is a measure of how the potentially creative design compares to other designs in its class in utility, performance, or attractiveness. Often this is a measure of how the design is valued by the domain experts or users and is either a weighted sum of performance attributes or is a reflection of the popularity of the artifact. To distinguish this from novelty, value is a measure of the design's performance rather than a measure of how the design's description differs from other designs 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. A predefined function of weighted value attributes is not appropriate because often a creative design can change the value system by introducing a performance or function that did not exist in the class of known designs before the creative design. For example, the Bloom laptop introduced a new performance measure for laptop designs: time to disassemble. Previously, laptops described their performance on environmental issues in terms of the type of materials used and their energy efficiency, not the amount if time to disassemble. A model for determining the value of a potentially creative design can adapt to new performance features if is based on the distance in performance criteria space from other artifacts, again, a measure that represents similar but different.

Surprise 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 recognizing the expected next difference. The amount of difference is not relevant as it is in the novelty metric, the variation from expectation is relevant. One way to think about measuring surprise is to characterize the existing designs in the design space as a probability distribution and determine the probability of the collection of attributes of the new design. The Bloom laptop introduced new description and performance attributes that were not considered design features in previous laptops. We have anecdotal evidence that some of the features were surprising, such as the importance of a removable keyboard.

## MEASURING NOVELTY, VALUE, AND SURPRISE

An assumption is made that a design can be described as a set of attribute-value pairs. For example, the conceptual space for the Bloom laptop design is the space of laptop or notebook computers. While the conceptual space need not be predefined or bounded, a list of attribute-value pairs that characterize this class of designs include technical specifications and performance features. Table 1 shows the attribute-value pairs that are used to describe the Apple Macbook, Macbook Air, and Macbook Pro designs. The last column in Table 1 shows the attribute values for the Bloom laptop design. Since the Bloom laptop is at the prototype stage, we have only included values for the attributes that are available for the prototype.

A design may have a structured description 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 design cannot be described as attribute-value pairs or decomposed into discrete parts. The clustering algorithms described below can be reformulated for other ways of representing or describing designs.

A formalization of creativity starts with a space of possibilities and an artifact within that space that is the product of 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 designs in that class. A subset of the class of artifacts, A, includes the known set of designs.

$$A = \{a_1, a_2, \ldots, a_n\} \ . \tag{1}$$

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

The evaluation, $E$, is a function of $a_i$.

$$E(a_i) = f(N(a_i), V(a_i), S(a_i)). \tag{2}$$

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 surprise of $a_i$

A principle for recognizing when a potentially creative design is creative is determining when the artifact is similar but different. In order for the artifact to be recognized and associated with a class of artifacts, it first must be similar to other artifacts. Once the similarity is established, the artifact is creative if it is different. We model "similar but different" in an artifact space using incremental and adaptive conceptual clustering so that new artifacts change the conceptual space over time rather than using a fixed measure of similarity. The distance function determines how far the potentially creative artifact is from the centroid of the nearest cluster of artifacts in the conceptual space. This allows us to treat the distance as a measure of the potentially creative artifact as similar (closest centroid) but different (distance from the center).

### Evaluating novelty: N

There are many accounts of measuring novelty using computational approaches. Marsland et al. [16] used Stanley's model of habituation [20] to implement a real-time novelty detector for mobile robots. Like the Kohonen Novelty Filter [12], 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. Saunders and Gero [21] 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.

We propose a model for measuring the novelty of a potentially creative artifact as a measure of the distance, d, between the centroid of the nearest cluster of the sets of description attributes of other artifacts in the space and the potentially creative artifact. For the laptop design example, the description space is defined by the technical specifications of a similar set of designs. In Table 1 we list the technical specifications of Apple Mac notebooks. This set of designs can be expanded to include other laptop designs: Toshiba, Sony, etc. We have only shown the Apple notebooks products because they demonstrate the nearest cluster of designs in the conceptual space. The Bloom laptop design is similar across all description attributes except the description of the body, the keyboard, and the touchpad. The Bloom laptop design has a modular body and removable keyboard and touchpad.

**Evaluating Value: V**

The value of a potentially creative artifact is a social phenomenon and determined by the "gatekeepers" as described by Csikszentmihalyi [7]. The value of any artifact is judged by criteria that are established by the requirements and performance attributes associated with the class of artifacts. Typically, value is determined using a weighted sum of the values of all requirements and performance attributes. Since a creative artifact can change our value systems, a potentially creative artifact can change the performance attributes for a conceptual space. Therefore, a predefined weighted sum function for determining the value of a potentially creative artifact is insufficient.

We propose a model for measuring the value of a potentially creative artifact using the same "similar but different" principle, that is, the value needs to be similar to others in its class, but may also be different by introducing new performance features or new types of values. Therefore, we measure value in the space of performance attributes of existing artifacts. The performance attributes are derived from the description attributes and/or represent social values of existing artifacts.

For measuring the value of a potentially creative artifact, we characterize the artifacts in the conceptual space in terms or performance attributes, and allow the potentially creative artifact to introduce new performance criteria. In this way, a distance measure to the nearest cluster of artifacts in the performance space characterizes the potentially creative artifact as similar but different. More specifically, value is a measure of the distance between the nearest centroid of the sets of performance attributes of the other artifacts in the space and the potentially creative artifact.

For the laptop design example, the Bloom design introduces a new performance attribute: disassembly time. The other performance attributes are similar to the Mac products, so this design is similar but different in value.

**Evaluating Surprise: S**

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. We can think of evaluating novelty and value as a slice in time, where the current time slice includes the attributes and values seen up until now. When a new artifact is introduced, that slice in time is updated, new attributes and/or values may be added, and new clusters of artifacts are formed. Surprise is recognized when we consider multiple slices of time. This is illustrated in Figure 2, showing how $A_1$ is surprising because it departs from expectations in the novelty and value space.

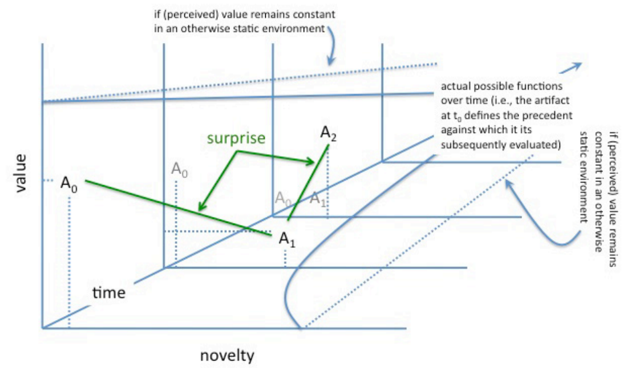


Figure 2. Evaluating surprise across time slices.

Horvitz et al [10] develop a model of surprise for traffic forecasting. The data used in this model was collected over 2 years and comprises traffic status in sensed traffic cells in Seattle, incident report data, contextual data such as holidays and weather. They generated a set of probabilistic dependencies among a set of random variables, for example linking weather to traffic status. When modeling surprise, they assume a user model that states that when an event occurs that has less than 2% probability of occurring, it is marked as surprising. They use a marignal model of the data, grouping incidents into 15 minute intervals. Surprising events in the past are collected in a case library of surprises. This provides the data for forecasting surprises based on current traffic conditions.

Itti and Baldi [11] describe a model of surprise for observing surprising features in image data using a priori and posterior probabilities. Given a user dependent model M of some data, there is a P(M) describing the probability distribution. P(M|D) is the probability distribution after the data is added, using Bayesian probability. Surprise is modeled as the distance d between the prior, P(M), and posterior P(M|D) probabilities.

Ranasinghe and Shen [19] develop a model of surprise as an integral part of surprise-based learning for developmental robots. In this model, surprise is used to set goals for learning in an unknown environment. The world is modeled as a set of rules, where each rule has the form:

Condition → Action → Predictions

A condition is modeled as:

Feature → Operator → Value

For example, a condition can be feature1 > value1 where "greater than" is the operator.

A prediction is modeled as

Feature → Operator

For example, a prediction can be "feature1 >" where it is expected that feature1 will increase after the action is performed. The comparison operators provided for surprise analysis include operators to detect the presence (%) or absence (~) of a feature, and the change in the size of a feature (<, <=, =, >=, >). If an observed feature does not match the prediction for the feature, for example, the feature was expected to increase and it decreased, then the system recognizes surprise.

The three models provide different approaches to modeling surprise based on the needs of the context in which they are developed. The Horvitz et al model [10] determines that an event in the past is surprising, and then it is a model for future surprising events. In the Itti and Baldi model [11], the new data is assimilated into the probability distribution, so something is surprising the first time it is introduced. The Ranasinghe and Shen model [19] do not use probabilities and instead find the first unexpected feature based on predictions of the direction in which the value of features will change.

For the laptop design example, the Bloom design introduces unexpected descriptions of two attributes, highlighted in Table 1. The body is modular when the trend in laptop design has been unibody. The keyboard and touchpad are removable when the trend has been fixed. Additionally, the Bloom design introduces two new performance features. In the Ranasinghe and Shen model of surprise this would be represented using the % (presence) operator to notice a feature that was not expected. The disassembly time is 2 min when that feature is not included in other laptop design descriptions (the Bloom team disassembled a several laptops and say the average time is 45 min). The removable keyboard, illustrated in Figure 3, was an emergent performance feature recognized by potential users as a valuable feature during the evaluation of the prototype. The design team included the removable keyboard to satisfy the modular design requirements and therefore it is part of the technical specifications, as well as a performance feature.

Figure 3. Removable keyboard in Bloom design became a performance attribute (Bohbe 2010)

**COMBINING NOVELTY, VALUE, and SURPRISE**

Combining novelty, value, and surprise is customized to an individual or group by assigning different weights to each of the characteristics. Figure 4 shows how the characteristics of creativity form a three-dimensional space. Artifacts in this space, including existing designs and a potentially creative design, can be compared visually by finding a surface that forms a subspace for potential creativity. This approach to combining novelty, value and surprise allows us to connect the formal models with the bias that different individuals or domains may have on evaluating creativity. For example, while the value of a potentially creative design may not be significantly different to existing designs, the effect of surprise may increase an individual's perception of the creativity of the artifact.

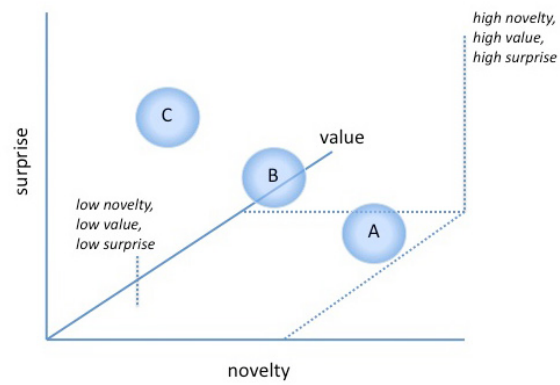


**Figure 4**. Combining Novelty, Value and Surprise

**NEXT STEPS: IMPLEMENTING THE EVALUATION METRIC**

In order to implement the evaluation as a computational system, we start with conceptual clustering to structure the design space for measuring novelty and value. When structuring the space for measuring novelty, we include descriptive attributes of existing designs. When structuring the space for measuring value, we include performance attributes of existing designs. Conceptual clustering allows us to characterize the designs in terms of their proximity to other designs by automatically grouping the designs into clusters. Once we have clusters, we can reason about similarities and differences.

Different clustering algorithms make different assumptions about the structure of the design space and for measuring the distance from the potentially creative design to a group or cluster of existing designs. The choice of a clustering algorithm depends on the characteristics of the design space, such as number of designs in the space, the number of attributes to describe the design, the density of different regions of the space. We describe two approaches to clustering to provide a sense of how the clustering can be implemented: K-means clustering (the algorithm was first published by Lloyd [13]) and Self-Organizing Maps (SOM) [12]. The distance measure for each approach to clustering is also defined.

K-means clustering uses a set of centroids to represent clusters of input data, or in our case, clusters of existing designs. 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:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{a_j \in S_i} \left\| a_j - \mu_i \right\|^2 \quad \textbf{(3)}$$

When K-means clustering is used to determine the distance of a potentially creative design, the update function is used to determine how far the new design 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 design where d is calculated using the K-means distance function:

$$d(a_i) = \sqrt{\sum \sum_{i=1}^{d} \Delta(k_i, a_i)^2} \quad \textbf{(4)}$$

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. The 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 design 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:

$$d(a_i) = \sqrt{\sum_{i} \Delta(k_i, a_{s_i})^2} \quad \textbf{(5)}$$

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 the distance to the nearest cluster of artifacts.

In addition to using clustering algorithms to structure the design space, we also use Bayesian probability to characterize the design space in terms of a probability distribution. Prior to the introduction of the potentially creative design, we have a set of known designs in a design space. We can express the prior probability of a design, D, in this space as P(D). Using Bayes Theorem, we can calculate the likelihood of the new design, H, as P(H|D). If the probability is less than a specified threshold, then we can say that the new design is surprising.

Our next steps are to use the data in Table 1 as a starting point for a set of existing designs in a design space of laptop designs. We have only shown Apple laptops to illustrate the attribute-value pairs of a cluster of designs that is similar to the Bloom laptop design. We will augment this list with other laptop design specifications. Using the clustering algorithms and probability distributions, we can plot the Bloom laptop design in the 3-dimensional space shown in Figure 4. We can compare the Bloom laptop to other designs in the space to visualize the relative creativity. We can also place the Bloom laptop design in the 3-dimensonal space with different weights for each of the three evaluation criteria: novelty, value, and surprise to visualize the relative creativity with different biases and preferences.

## CONCLUSIONS

This paper argues for an approach to evaluating creative designs that is independent of the design discipline and of the source of creativity. Our approach uses AI models that operate in the conceptual space of the design discipline, thereby contextualizing the evaluation and providing a relative measure of creativity rather than a binary judgment. Formalizing the essential criteria for evaluating creativity allows us to compare the many different approaches to developing computational systems that are themselves creative as well as computational systems that enhance human creativity. With such a metric, we have a common ground for evaluating creativity in human, computer, and collectively intelligent systems.

The three essential criteria for evaluating creativity are novelty, value and surprise. 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 a cluster of similar, known artifacts. 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 satisfy domain specific performance criteria and possibly extend our understanding in a specific field, change our value system, or enhance our lives in some way. Measuring value is also based on a distance metric, showing how the value of a creative design is similar but different from the value of clusters of existing designs. Surprise is an aspect of creativity that we recognize when we say that something is creative because it does not meet our expectations for the next design in its class. Surprise is measured using probability functions that can identify when one or a set of features is not expected, or by prediction rules that can identify when a specific feature was not predicted.

The contribution of this paper is the articulation of three characteristics of creativity that can be used as a basis for evaluating a potentially creative design. The paper shows how a common metric is derived from the various definitions and metrics developed in different disciplines. The elements of the metric are not new, but the combination of these three characteristics is presented as a common model for evaluating creativity. The metrics are developed further using various AI techniques that can be adapted and applied to different contexts and conceptual spaces as a computational approach to evaluating creativity or as a guide for human judgment of creativity.

## REFERENCES


1. Amabile, T. Social Psychology of Creativity: A Consensual Assessment Technique. *Journal of Personality and Social Psychology*, 1982, **43**(5):997-1013.
2. Andreasen, N. *The creating brain*. New York: Dana Press, 2005.
3. Bailin, A. *Achieving extraordinary ends: An essay on creativity*. Boston, MA:Kluwer Academic, 1988.
4. Bean, R. *How to develop your children's creativity*. Los Angeles, CA: Price Stern Sloan Inc. 1992.
5. Bhobe, R., Engel-Hall, A., Gail, K., Huotari, J., Koskela, M., Liukas, L. and Song, C.: *Bloom: Mechanical Engineering 310 Spring Design Proposal*, Autodesk, 2010. http://students.autodesk.com/?nd=sustainable_standard&material_id=106&course_id=15
6. Boden, M. *The Creative Mind: Myths and Mechanisms*, Routledge; 2nd edition, 2003.
7. Csikszentmihalyi, M. *Creativity: Flow and the Psychology of Discovery and Invention*, HarperCollins Publishers, 1996.
8. Csikszentmihalyi, M. and Wolfe, R. New Conceptions and Research Approaches to Creativity: Implications of a Systems Perspective for Creativity in Education, in Kurt Heller, *The International Handbook of Giftedness and Talent 2nd edition*, Elsevier, 2000, pp 81-94.
9. Gero, J.S. Computational Models of Innovative and Creative Design Processes, *Technological Forecasting and Social Change*, 2000, **64**:183-196.
10. Horvitz, E., Apacible, J., Sarin, R. and Liao, L. Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service, *Proceedings of the Conference on Uncertainty and Artificial Intelligence 2005*, AUAI Press, July 2005.
11. Itti L. and Baldi P. A Surprising Theory of Attention, *IEEE Workshop on Applied Imagery and Pattern Recognition*, Oct 2004.
12. Kohonen, T. *Self-Organisation and Associative Memory*, Springer, Berlin, 1993.
13. Lloyd, S.P. Least Squares Quantization in PCM, *IEEE Transactions on Information Theory,* 1982, **28**(2): 129–137.
14. Maher, M.L. Evaluating Creativity in Humans, Computers, and Collectively Intelligent Systems, *DESIRE'10: Creativity and Innovation in Design*, Aurhus, Denmark, 2010.
15. Maher, M.L., Boulanger, S., Poon, J., and Gomez de Silva Garza, A. Exploration and Transformation in Computational Methods for Creative Design Processes, in J. S. Gero, M. L. Maher and F. Sudweeks (eds), *Preprints Computational Models of Creative Design*, University of Sydney, 1995.
16. 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.
17. Mumford, M. D. Where have we been, where are we going? Taking stock in creativity research, *Creativity Research Journal*, 2003, **15**:107-120.
18. Oman, S and Tumer, I. The Potential of Creativity Metrics for Mechanical Engineering Concept Design, in Norell Bergendahl, M., Grimheden, M., Leifer, L., Skogstad, P., Lindemann, U. (Eds) *Proceedings of the 17th International Conference on Engineering Design (ICED'09),* Vol. 2, pp 145-156, 2009.
19. Ranasinghe, N. and Shen, W-M. Surprise-Based Learning for Developmental Robotics. In *Proc. 2008 ECSIS Symposium on Learning and Adaptive Behaviors for Robotic Systems*, Edinburgh, Scotland, August 2008.
20. Runco, M. A. *Creativity: Theories and Themes: Research, Development and Practice*. Amsterdam: Elsevier, 2007.
21. Saunders, R., Gero, J.S. Designing For Interest and Novelty: Motivating Design Agents, *CAAD Futures 2001*, Kluwer, Dordrecht, pp 725-738, 2001.
22. Shah J., Smith S., Vargas-Hernandez N. Metrics for measuring ideation effectiveness, *Design Studies*, 2003, **24**(2):111-134.
23. Solomon, B., Powell, K., and Gardner, H. Multiple Intelligences. In Runco, M.A. and Pritzker, S. (Eds), *Encyclopedia of Creativity*, San Diego, CA:Academic Press, pp 259-273, 1999.
24. Stanley, J.C. Computer Simulation of a Model of Habituation. *Nature*, 1976, **261**:146-148.
25. Villalba, E. On Creativity: *Towards an Understanding of Creativity and its Measurements*, Luxembourg: Office for Official Publications of the European Communities, Joint Research Centre, Report JRC 48604, 2008.
26. Wiggins, G. A Preliminary Framework for Description, Analysis and Comparison of Creative Systems, *Knowledge-Based Systems*, 2006, 19:449-458.

**Table 1. Laptop Design Technical Specifications and Performance Features**

**Technical specifications**

| | MacBook | 11-inch MacBook Air | 13-inch MacBook Air | 13-inch MacBook Pro | 15-inch MacBook Pro | 17-inch MacBook Pro | Bloom |
|---|---|---|---|---|---|---|---|
| **Body** | Polycarbonate unibody | Precision aluminum unibody | Precision aluminum unibody | Precision aluminum unibody | Precision aluminum unibody | Precision aluminum unibody | Modular components |
| **Processor** | 2.4GHz Intel Core 2 Duo | Up to 1.6GHz Intel Core 2 Duo | Up to 2.13GHz Intel Core 2 Duo | Up to 2.7GHz dual-core Intel Core i7 processor | Up to 2.3GHz quad-core Intel Core i7 processor | Up to 2.3GHz quad-core Intel Core i7 processor | 2.4GHz Intel Core 2 Duo |
| **Height** | 1.08 inch | 0.11 to 0.68 inch thin | 0.11 to 0.68 inch thin | 0.95 inches | 0.95 inches | 0.95 inches | 1.08 inch |
| **Display** | 13.3-inch LED-backlit | 11.6-inch LED-backlit | 13.3-inch LED-backlit | 13.3-inch LED-backlit | 15.4-inch LED-backlit | 17-inch LED-backlit | 13.3-inch LED-backlit |
| **Trackpad** | Multi-Touch trackpad | Multi-Touch trackpad | Multi-Touch trackpad | Multi-Touch trackpad | Multi-Touch trackpad | Multi-Touch trackpad | Removable Multi-Touch trackpad |
| **Processor** | 2.4GHz Intel Core 2 Duo processor | 1.4GHz or 1.6GHz Intel Core 2 Duo processor | 1.86GHz or 2.13GHz Intel Core 2 Duo | 2.3GHz dual-core Intel Core i5 or 2.7GHz dual-core Intel Core i7 processor | Up to 2.3GHz quad-core Intel Core i7 processor | Up to 2.3GHz quad-core Intel Core i7 processor | 2.4GHz Intel Core 2 Duo processor |
| **Memory** | 2GB or 4GB memory | 2GB or 4GB memory | 2GB or 4GB memory | 4GB or 8GB memory | 4GB or 8GB memory | 4GB or 8GB memory | 2GB or 4GB memory |
| **Storage** | Up to 500GB 5400-rpm hard drive | Up to 128GB flash storage | Up to 256GB flash storage | Up to 500GB 5400-rpm hard drive | Up to 750GB 5400-rpm hard drive | Up to 500GB 7200-rpm hard drive | 256MB of DDR3 SDRAM |
| **Graphics** | NVIDIA GeForce 320M graphics processor | NVIDIA GeForce 320M graphics processor | NVIDIA GeForce 320M graphics processor | Intel HD Graphics 3000 | Intel HD Graphics 3000 | Intel HD Graphics 3000 | NVIDIA GeForce 320M graphics processor |
| **Display** | 13.3-inch LED-backlit | 11.6-inch LED-backlit | 13.3-inch LED-backlit | 13.3-inch LED-backlit | 15.4-inch LED-backlit | 17-inch LED-backlit | 13.3-inch LED-backlit |
| **Resolution** | 1280 by 800 pixels | 1366 by 768 pixels | 1440 by 900 pixels | 1280 by 800 pixels | 1440 by 900 pixels | 1920 by 1200 pixels | 1280 by 800 pixels |
| **USB ports** | Two USB 2.0 ports | Two USB 2.0 ports | Two USB 2.0 ports | Two USB 2.0 ports | Two USB 2.0 ports | Three USB 2.0 ports | Two USB 2.0 ports |
| **Camera** | iSight camera | FaceTime camera | FaceTime camera | FaceTime HD camera | FaceTime HD camera | FaceTime HD camera | iSight camera |
| **Keyboard** | Full-size keyboard | Full-size keyboard | Full-size keyboard | Full-size, backlit keyboard | Full-size, backlit keyboard | Full-size, backlit keyboard | Removable full-size keyboard |

**Performance features**

| | MacBook | 11-inch MacBook Air | 13-inch MacBook Air | 13-inch MacBook Pro | 15-inch MacBook Pro | 17-inch MacBook Pro | Bloom Laptop Design |
|---|---|---|---|---|---|---|---|
| **Disassembly time** | | | | | | | 2 min |
| **Removable keyboard** | | | | | | | Yes |
| **Price** | Just $999 | From $999 | From $1299 | From $1199 | From $1799 | From $2499 | From $1000 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Weight** | 4.7 lbs | 2.3 lbs | 2.9 lbs | 4.5 lbs | 5.6 lbs | 6.6 lbs | 4.7 lbs |
| **Battery life** | 7 hours | 5 hours | 7 hours | 7 hours | 7 hours | 7 hours | 7 hours |
| **Recyclable Materials** | Recyclable polycarbonate enclosure | Highly recyclable aluminum and glass enclosure | Highly recyclable aluminum and glass enclosure | Highly recyclable aluminum and glass enclosure | Highly recyclable aluminum and glass enclosure | Highly recyclable aluminum and glass enclosure | Highly recyclable plastic and glass enclosure |
| **Non-toxic materials** | Mercury-free LED-backlit display | Mercury-free LED-backlit display | Mercury-free LED-backlit display | Mercury-free LED-backlit display | Mercury-free LED-backlit display | Mercury-free LED-backlit display | Mercury-free LED-backlit display |
| | Arsenic-free display glass | Arsenic-free display glass | Arsenic-free display glass | Arsenic-free display glass | Arsenic-free display glass | Arsenic-free display glass | Arsenic-free display glass |
| | BFR-free | BFR-free | BFR-free | BFR-free | BFR-free | BFR-free | BFR-free |
| | PVC-free6 | PVC-free3 | PVC-free3 | PVC-free3 | PVC-free3 | PVC-free3 | PVC-free6 |
| **Energy efficiency** | Meets ENERGY STAR Version 5.0 requirements | Meets ENERGY STAR Version 5.2 requirements | Meets ENERGY STAR Version 5.2 requirements | Meets ENERGY STAR Version 5.2 requirements | Meets ENERGY STAR Version 5.2 requirements | Meets ENERGY STAR Version 5.2 requirements | Meets ENERGY STAR Version 5.0 requirements |
| | Rated EPEAT Gold7 | Rated EPEAT Gold4 | Rated EPEAT Gold4 | Rated EPEAT Gold4 | Rated EPEAT Gold4 | Rated EPEAT Gold4 | Rated EPEAT Gold7 |