A Brief Review of Approaches to Design Novelty Assessment

David C. Brown
Worcester Polytechnic Institute, Worcester, MA, USA

This paper questions the assumption in design creativity research that we know how to calculate a meaningful score for the novelty of a design. This paper looks at the basics of novelty calculation, and reviews a variety of approaches, attempting to provide a rough categorization of them. Novelty is a key ingredient in the evaluation of the creativity of a product. Each ingredient of any assessment of creativity needs to be calculated as carefully and accurately as possible, so novelty assessment is very important.

{Please cite this paper as WPI Computer Science TR-21-03}

1. Introduction

A well-hidden assumption in design creativity research is that we know how to calculate a meaningful score for the novelty of a designed artifact. By disseminating information about different approaches in creativity research to assessing novelty we hope to counter the tendency to restrict the methods actually used, and help design creativity research make progress.

The paper focuses on actual or potential *computational* methods. We start by examining novelty and its definitions, questioning the role of "difference". Next the use of novelty in creativity theory is presented, including how it is done for computer generated stories: a special kind of designed artifact. Focus moves to examine the main variations of novelty assessment schemes that might be used when evaluating creativity. For another point of view, we briefly introduce views of similarity from Cognitive Science, as similarity is the basis for most novelty calculations. The paper finishes with an examination of some of the assumptions around novelty assessment, and a conclusion.

1.1 A Note about this Paper

This paper is based on one written during 2015-17 that was not published. A significant collection of relevant new references about novelty, originality, and similarity have been added to the end of the paper. They are not cited in this text. Interesting new references to begin with include Barto et al. [2013], Hay et al. [2019], Lamb et al. [2018], and Siddarth et al. [2020].

Note that some of the many older and often cited references are not included here, but can be easily found in the reference list of papers such as Chakrabarti & Khadilkar [2003], Brown [2014b] and Jagtap [2018].

1.2 The Role of Novelty in Creativity

Novelty is seen by almost all researchers as the key ingredient in the evaluation of the creativity of a product (e.g., [Besemer 2006; Pease et al. 2001; Ritchie 2007]). Here we focus on the product and not the process [Jordanous 2016]. As each ingredient of any assessment of creativity needs to be made as carefully and accurately as possible, an understanding of novelty assessment is very important.

There are many different ways that novelty is assessed in the research literature. This paper does not aim to suggest which as more successful, accurate or appropriate. However, an understanding of the many variations of novelty evaluation should help improve design creativity research. We hope that this paper provides inspiration for further study of novelty in cognitive psychology, marketing, computational design, and computational artistic creativity.

The US Patent Office has long been concerned with novelty, stipulating that inventions should be *new*, useful and nonobvious. But these criteria are seen as qualitative. For computation, we require that no human inferences/assessment be explicitly included. Researchers concerned with patents have fairly recently attempted to see "new" in a quantitative light. Simonton [2012], for example, proposes a quantitative value for novelty based on the probability of a design's generation: but he provides no method for novelty calculation.

2. The Basic Ingredients of Novelty Evaluation

Here we refer to a subset of ingredients for evaluating creativity [Brown 2014a]. These ingredients are:

- a *description* of the complete or partial artifact being judged;
- the *agent* judging (i.e., person, computer program, or group);
- the *temporal* basis for comparison
(e.g., the point in time or the time period);
- the *source* of the set of designs used as the basis for comparison (e.g., personal, group, industry, global);
- the *method* of evaluation for novelty.

Variations in these ingredients affect the result of any evaluation. Note that novelty is a "moving target": in response to any new design being added to the basis for comparison, to the judge having access to new information, and to the effects of time on the judgment. The temporal basis for comparison is usually grounded at assessment time or at design completion.

By *design* we mean a description of a design; an *artifact* means a thing. We are not concerned with a different ‘instance’ of a designed artifact (such as multiple instances of a particular type of phone in a shop) but differences between descriptions. To judge novel functionality/use we may need to have the artifact available.

We assume use of some standard, normalized description: allowing designs to be compared; avoiding the difficult problem of having the same artifact being described with a differently formed description.

Note that many authors allow an *attribute* to be almost anything in the description: including parameters such as “length”, component types such as “gear”, abstractions such as function, but also relationships such as “touching” and derived performance parameters such as the amount of “heat produced”.

Grace et al. [2015] write about references to the “whole” (e.g., iPhone4) or to attributes (e.g., storage capacity): statements about novelty might involve either. Referring to a type (e.g., a “phablet”) can also restrict the descriptions used as a basis for comparison. Other potentially relevant abstractions might be based on function, behavior or structure.

The “source” refers to from where the basis for comparison is gathered. The basis might only be designs produced by one designer. This corresponds to evaluating for P-Creative designs [Boden 1994]. By using a global source, and by using “all history” as the temporal basis, we are evaluating for H-Creative designs. By varying the size of the source-time space, different comparisons are possible. By focusing just on novelty, we can identify *H-novel* and *P-novel* designs.

Dean et al. [2006] note that the novelty of any idea is judged in relation to how uncommon it is *in the mind of the idea rater* or how uncommon it is *in the overall population of ideas*: i.e., it depends on the “judge” and the source-time scope of their comparison set. This allows a design to be P-novel for a user but not necessarily P-novel for the designer.

3. Examining the definitions of novelty

The meaning of “novelty” is complicated. By examining a variety of dictionary definitions of novelty we find that most are some variation on “the quality of being *new*, *original*, or *unusual*”. These terms reveal definitions such as:

> *New*: “Not existing before; made, introduced, or discovered recently or now for the first time”.
> *Original*: “Not dependent on other people’s ideas; inventive”.
> *Unusual*: “Not habitually or commonly occurring or done”.
> *Nonobvious*: “at an adequate distance beyond or above the state of the art”.

The underlying concepts are: being completely *new* (i.e., occurring for the first time); being *inventive* (i.e., including unusual components, structure or behavior); *distance* from existing designs (i.e., concerning the degree of similarity); *frequency* of occurrence of a design in a set of existing designs.

Dean et al. [2006] refer to three novelty-related constructs: *rarity*, *originality*, and *paradigm relatedness*.

Rarity: the rarity of an idea is how uncommon (infrequent) an idea is in a set of ideas. So "New" implies "unique" (not existing before), while "Rare" implies only a few examples are known (uncommon). Note that it requires some kind of (subjective) *similarity determination* (i.e., distance). An important issue is that this allows only *slightly different* artifacts in a small comparison set to be considered "unique" or "rare". Rarity depends on comparison, and hence the scope of the comparison set.

Originality: artifacts have originality if they are i) rare, but ii) also "have the characteristic of being ingenious". Detection of ingenuity implies recognizing the use of very different structure, behavior or sub-functions in order to achieve the same overall functionality (i.e., inventive). This implies having a structured design representation, as well as some sort of summary/frequency knowledge, or type clustering.

Paradigm Relatedness (PR): this is concerned with whether the artifact is *germinal* or *transformational*. It is the degree to which an idea maintains or modifies a paradigm.

PR Germinal: this is about "trend setting" or "influential" ideas. These designs contribute to changing the paradigm. It seems to require a temporal record of the development of artifacts over time, and whether one artifact was influenced by another. A scale for this might involve the number of designs influenced and for how long that influence persisted, although it is unclear exactly how such things might be measured. Note that it has no utility for new designs, only for past designs.

PR Transformational: This is about 'how' the originality was produced: the types of changes made to prior artifacts. These changes may involve the elements in the design (same or different), or the relationships between elements (same or different). The category "paradigm modifying" requires at least one of these changes; "transform" has both different; "extend" has the elements different; "redesign" has the relationships different. This provides a crude explanatory scale for the amount of change, and could be applied to a new design relative to a comparison set. Note that they all require differences to be detected and the design to be described in terms of elements and relationships (i.e., a basic structured representation).

It is clear that Dean et al.'s novelty-related constructs are not completely independent. We conjecture that the more germinal a design is, the more likely it is to be transformational: the more germinal it is, the more likely it is to be seen as ingenious and therefore more likely to be seen as original.

In summary, the four novelty-related concepts, *new*, *inventive*, *distance* and *frequency*, need to be augmented with a fifth, *paradigm changing*, in order to cover the main concepts related to novelty assessment.

4. Examining Difference

Definitions of the term "new" revolve around the dimensions "recent" or "different". Our initial, simplifying view is that "new" means different from *all* previous designs: there is not complete similarity, although there could be a high similarity score. In what follows, the point is to indicate some of the basic subtleties, not to be exhaustive.

We assume that a comparison set of previous designs (D) is available as the basis for comparison with the design being evaluated. This collection might be organized and indexed in a variety of ways in order to suit the newness calculations being made. Possibilities include ordered by time, indexed by attributes, and clustered around stereotypes, etc.

Grace & Maher [2014] point out that such comparisons with descriptive representations are based on what they call the "expectation of continuity": that is, future designs will conform to the representations already learned. Here we are just concerned with D's role as a basis for comparison. Note that the judging agent might be the designer, user, or a computer, where each judge might have different knowledge, and therefore different results.

4.1 Definitions of Difference

The simplest, most primitive meaning of "difference" is "not *exactly* the same". For example, considering single attributes of design d:

- A design with color = "blue" versus color = "red";
- A design with color = "scarlet" versus color = "red";
- A design with length = "3" versus length = "7";
- A design with length = "3" versus length = "3.1";
- A design with length = "3" versus length = "3.0000000000001".

As the five alternatives did not previously exist then, strictly, they are all new. We will refer to that as Absolutely New (absnew).

1. absnew(d) is true if the value of at least one attribute of design d is not exactly the same as its matching attribute when comparing d to every design from the set of existing designs (i.e., the set D).

This covers the five cases above (i.e., color change; length change). If there is at least one design in D that has all matching attributes with exactly the same values, then absnew is false. Note that if there are only very few that are the same, and the size of D is large, then there is an opportunity for a novelty definition that involves rarity.

2. absnew(d) is true if there are no designs in D that have exactly the same set of attributes.

Note that this covers the first design of its type (e.g., the first automobile), as well as types with additional characteristics.

Restrictions on D:

For a practical model of the judgment of newness it makes no sense to include *all* existing designs as the target of comparison for difference. Consequently, we suggest some restrictions, with combinations possible.

a) Do not compare different types (e.g., milk jug, car).
b) Do not call relatively small changes "different".
c) Do not compare things at different levels of abstraction.
d) Use the time of introduction of the artifact being considered.
e) Restrict the comparison set in other ways.

We use Dr to refer to the restricted set of designs, regardless of which restrictions have been applied, and absnewDr to refer to the test of newness in this new setting.

Restriction a: let Dr be the subset of D that contains "comparable" design descriptions that refer to the same general type of artifact. We assume the existence of a taxonomy, or some other organizing knowledge structure. If there are no comparable design descriptions then absnew is true, as this design introduces a new type.

Restriction b: let "different" be relaxed from "not exactly equal" by defining a decision function for *every* type of attribute, in the context of Dr.

e.g., different(d, length, 1", 1.001", Dr) is false;
different(d, color, red, scarlet, Dr) is false.

This might be done by defining a threshold for the difference in the values of a length, such that a difference of 0.5 gives true while a difference of 0.001 gives false. One can imagine this threshold for a meaningful difference being context dependent and changeable.

For absnew *only one* different predicate needs to be true. An extension that is easy to imagine is a "fuzzy" decision function that provides a degree of difference, instead of a binary result.

Restriction c: let D be refined to Dr by restricting the subset used for comparison to those design descriptions at the same level of abstraction.

This restriction needs to be handled with care, as for example "red" might match "warm colored". However, this is a semantic judgment that is a more advanced kind of matching. Note that this restriction is important for maintaining something close to an absnew match when considering intermediate or conceptual designs. Note too that things that might be considered new with a description at one level of abstraction might not be at another: e.g., any detailed design description of a dramatically new building might be reduced to "a roof with walls" after enough abstraction.

Restriction d: let all comparisons be made with the set of design descriptions Dr selected to represent an appropriate time or period.

The Sydney Opera House at the time of its introduction would have been considered to be new (i.e., different enough), but in today's context it would not be, as the comparison set has changed. By setting Dr at a particular time (Dtr) it becomes possible to ask hypothetical questions: e.g., if design d had been introduced at time t would it have been considered new?

Restriction e: other ways to restrict D to an appropriate subset might include characteristics of the designer, such as region of origin, level of expertise, reputation, age, etc.

Multidimensional boundaries:

These restrictions provide boundaries defining different sets of designs:

- The set of designs from D for which absnew is false (i.e., designs that are exactly the same as d).
- The set of designs from D for which absnew is true (i.e., designs that have at least one attribute's value not exactly the same).
- The set of designs for which absnew is true, and which fall inside Dr, for some set of restrictions on D (i.e., the designs are relevant, the differences meaningful, etc.).
- The set of designs for which absnew is true, but are outside of Dr, for some set of restrictions on D (i.e., the designs are either not relevant, the

differences not meaningful, etc.). If absnew is true, but absnewDr is false, then the newness of these designs is undefined, but they could be taken to be new if needed, as absnew is true.

4.2 Relaxing the match with key attributes

There may be attributes about which we "don't care", or are unimportant. This might modify absnew to "require" some attributes to be true, relax the newness of the others, or even ignore some attributes.

Many methods in the literature use weighted attributes. What is considered important might be conditional on some attribute or attributes [Grace & Maher 2014]. Importance might depend on the existing design being used for comparison: comparing a design d with a representative of designs of a certain type may set importance depending on that type.

4.3 Moving away from binary judgment

To compare the difference in newness between whole designs, and not just individual attributes, some form of "accumulated" difference needs to be formed. We will not discuss this further at this point, but a simple computational model might just sum the normalized differences. A full model of newness would need other accumulation methods to be included.

5. Novelty in Creativity Theory

Ritchie's [2007] criteria use a program's input and its output (i.e., a design), with measures that human judges might apply. Ritchie defines novelty as "To what extent is the produced item dissimilar to existing examples of its genre", with judgments of novelty based on "typicality" and "innovation". Typicality is defined as "To what extent is the produced item an example of the artefact class in question". Low typicality scores are due to being "dissimilar to the norm for that class". He defines innovation as "How different is the item from those that guided its original construction in the system". Innovation is an internal-based measure, while typicality is an external-based measure.

Any assessment of novelty external to the program probably does not use knowledge of the *Inspiring Set* (i.e., designs built into the system). Novel results may not be very typical, but new items should at least have acceptable typicality. Pruning highly typical or highly atypical output would be appropriate, although typicality may change over time.

Pease et al. [2001] point out that a judge's understanding of the knowledge that the creator has affects their judgment of the creator's creativity [Brown 2014a]. They assume that creative items must be (at least) novel, and that to be novel they must not be too similar to an existing item (boring), nor too different from all existing items (bizarre).

Pease adopts Boden's idea [1994] that an item is *not* novel if it is from inside a well explored area of a concept space; it is *merely* novel if it is from inside the portion of the space that is not yet explored; it is *fundamentally* novel if it is from outside the space but it is not bizarre; or it is *bizarrely* novel if it is from far outside the space. Object-level procedures O are directly used to generate or evaluate d; and meta-level procedures are used to generate or evaluate a member of O. Pease defines novelty as *fundamental* if d could not have been generated without meta-level procedures; as *merely* otherwise. It isn't clear how this can be used to assess novelty in a practical way.

6. Novelty Assessment for Stories

Some of the work in computational creativity concerns stories [Perez & Sharples 2004]. Peinado et al. [2010] review work on the assessment of novelty of computer-generated stories, where we consider a story to be a very special type of design. It requires comparisons to known examples using some form of similarity metric: novelty is inversely proportional to similarity given that Knowledge Base (KB). Note too that some designs can be described, and compared, using text.

Novelty assessment for stories uses the *structure* and specific *aspects* of stories to measure similarity. The aspects used for similarity are dependent on the fact they are stories with structure: not necessarily what the stories are 'about'. Different aspects have different importance (i.e., weights).

The Mexica system [Pérez & Sharples 2001] is unusual in that it assesses the novelty of the story *in progress*. They base this on the *similarity* of actions and their *frequencies* in the story in progress compared to a KB of previous stories (by humans). For Pérez et al. [2011] there are three characteristics of story novelty: sequence of actions; structure of the story; use of characters and actions. If a story is novel enough it is added to the KB. Simple distance measures are defined for each aspect of the story. One novelty score used is the ratio between the new knowledge which is different to what was previously stored, against the new knowledge (i.e., new-but-different / new).

The ProtoPropp system [Peinado & Gervás 2006] starts with a modified member of the *inspiring set* [Ritchie 2007]. The system uses "formal metrics" for novelty: systematic element-by-element comparison between a proposed story and the KB. Elements include characters, locations, etc. Different weights are used for different types of comparisons. The average value of the comparison scores gives an indication of the novelty. Management of that KB is important, as new stories added make it bigger (i.e., a *dynamic inspiring set*). Therefore, indexing or clustering is needed to reduce the complexity of novelty checking.

Peinado & Gervás [2006] did an empirical study with human readers, resulting in the construction of equations for novelty. The equations are based on similarity judgments: (e.g., of events). They determine that novelty of *events* is most important, followed by *characters*, with *props* and *scenarios*: enforced in their metrics with weights: e.g., for characters, a new character is considered more novel than a replacement with a different character type. *Story novelty* is the weighted sum of all the ingredient novelties. They suggest that metrics could be used to detect what might be changed to increase novelty: just as Besemer [2006] does for creativity.

Note that as novelty assessment of stories depends on the them being structured, and being stories: i.e., it could be an indication that accurate assessment of designs may be dependent on the type of the design.

7. Criteria for Aesthetic Functions

Computational creativity research usually targets things where Aesthetic judgements are needed, such music or poetry. Aesthetic functions can be used to assess all sorts of qualities, including novelty. Colton et al. [2015] propose metrics that allow comparisons between aesthetic functions. *Specificity* reflects whether an aesthetic function can provide a total order over a set of objects: i.e., it can distinguish between them. Clearly, high-specificity is preferable. *Agreement* measures the amount of agreement between two aesthetic functions: especially useful if they are related in some way, such as novelty and surprise. *Transitive consistency* looks for how consistently A < C, where A < B and B < C, and "<" means "is less aesthetically preferable". Those preferences can be "complex and multi-objective" and so there may not always be consistency. Similar points have been raised in the discussions of Cognitive Science approaches to similarity (see section 9).

8. Variations in Novelty Assessment Schemes

The common variations in novelty evaluation schemes can be organized by using the ingredients for evaluating creativity listed above, and by involving the five novelty-related concepts. As a reminder, these are the description, the agent, the temporal basis, the source, and the method.

8.1 Variations in the description of the designed artifact being judged

Product vs. process:

For a designed and manufactured product the user has access to the artifact, or pictures of it, allowing them to judge its novelty. However, there is usually very little information about, or indication of, the creative processes involved. Hence the user judges the artifact, not the creative process [Ritchie 2007]. In contrast, for artistic artifacts, the creative process may be visible in the marks left by the artistic processes (e.g., painting, sculpting). The judgment of artistic creativity relies much more on judging the development process with the novelty sometimes taken as given.

Treating design as a single level vs. a multi-level description:

Russell & Norvig [2009] refer to types of descriptions: *atomic*, no internal structure; *factored,* a vector of attributes and values; *structured*, with attributes and relationships. Treating the design as a single level factored description makes similarity/difference detection easier. However, despite being a harder to process, multi-level descriptions more accurately express design complexity. In addition, novelty at different levels may affect the overall judgment of novelty by different amounts.

There are a variety of types of multi-level descriptions used in novelty evaluation. These can be *functional* representations that explicitly include Function, Behavior and Structure (FBS) [Erden et al. 2008]; they might reflect design stages; or reflect the causality that allows a design to function (e.g., the SAPPhIRE model [Srinivasan & Chakrabarti 2010]).

Shah et al. [2003] proposed commonly adopted metrics for novelty in Engineering Design. They state that "novelty is a measure of how unusual or unexpected an idea is compared to other ideas". While for a different metric Shah refers to a structured representation (i.e., physical principles, working principles, embodiment & detail), these levels are not used for novelty: the design stages "conceptual" and "embodiment" are used instead. This biases the approach away from finished designs towards conceptual designs. For variations on Shah's approach see [Brown 2014b].

Shah's first novelty metric considers all the existing ideas for the given design problem. Then a set of "key" attributes (i.e., "functions or characteristics") is determined at the conceptual and/or the embodiment stage.

A novelty score, from a preset range, is assigned for each value found for each attribute: more *frequent* values get lower scores, while unusual values get higher scores, indicating more novelty. Note that finding the novelty score for a value may involve inexact matching. Each attribute (e.g., function) is given an "importance" weight. Each stage is also given a weight, as the conceptual stage probably makes more impact on novelty.

The second novelty metric, with more potential for automated calculation, is based on the set of ideas (a comparison set) generated by participants in a design experiment. It starts by identifying the key attributes of the ideas: typically functions. Then all the values are found that have been produced for those attributes. Next count how many instances of each value there are. This is done for both conceptual and embodiment stages.

The method produces a weighted score across all functions and stages for a design. Weights are used to represent the importance of each function, and the importance of each stage. The novelty score for a value for a single function at a single stage is based on ((T – C) / T) where T is the total number of values produced for the function in that stage, and C is the number of times the value from this design was used in the comparison set.

If the value is *rare*, C will be small and the score will be closer to 1. Deciding that a value belongs in the C count depends on inexact matching. As a *frequency*-based measure, relative to values produced by the participants, it seems to be a measure of the capabilities of those participants.

Dean et al.'s [2006] *transformational* view of *paradigm relatedness* rests on the types of changes made to prior artifacts. These changes may involve the elements in the design, and/or the relationships between elements. Lopez-Mesa & Vidal [2006] propose a bigger distinction, distinguishing between parts, structure, function and the whole design: each change being more paradigm-changing. Designers (or teams) are rated by how many of each type of change is produced. A second method, also using function, structure, and details, evaluates a team for *non-obvious* performance by comparing it with the performance of others. The more of the other teams with a similar solution at the same level, the more obvious it is. Note that these methods seem more about comparing the teams' ability to be novel, than they are about assessing the novelty of the designs.

Sarkar & Chakrabarti [2011] propose that the higher the level in the SAPPhIRE model at which there is a difference between d and existing designs, the more likely it is to be novel. The model includes actions at the top, changes of state below, and physical phenomena below that.

The model is mapped to the FBS model to allow 'human assessment' of the degree of novelty (1 to 4). If d's function does not already exist then it is "very high novelty". At the other end of the scale, if the structure is different, but the behavior is not different, this is "low novelty".

Prior to that approach, Srinivasan & Chakrabarti [2010] estimated the novelty of a concept by giving scores, between 1 and 7, that directly corresponded to the parts of the SAPPhIRE model.

Chakrabarti & Khadilkar [2003] proposed a more complicated model, although its predictions were worse than the 2011 model. They use a design model with the vertical levels: need, task, subsystem structure, working principle, technology, and implementation. The design d is compared with the reference and "differences" identified at each level. Higher levels have more impact on the novelty. At each level, horizontal weights are associated with the importance with respect to the contribution of that difference to the functionality (main function; supplementary; additional). The novelty value at each level is affected by the horizontal weights, and the results are aggregated until a single value for novelty is obtained.

A potential problem with the use of multi-level descriptions for novelty assessment is that comparison sets are most likely to be finished, detailed design descriptions, with no multi-level structure, and no design history available: getting such information could be difficult and expensive.

Another issue is the use of 'weights' to represent different degrees of novelty depending on the level of abstraction of the design description, with detailed part descriptions being the lowest. While intuitively promising, as functional differences have more impact than small, detailed differences, it builds 'guesses' of the degrees of novelty into the algorithm, based on coarse differences. However, there may be ways to represent and measure those differences. For example, Murakami & Koyanagi [2017] propose a 'delta' representation to record design differences, as well as a way to determine functional similarity.

8.2 Variations in the set of designs that act as the basis for comparison

Novelty evaluations may concern a single design, or a set of designs with a common source, such as the same designer or design group. For a set, the emphasis is more on establishing the general ability of the designer/group to be consistently creative. In general, the scope might be from a certain group, designers from the same company, the same industry, or some more global scope: i.e., less than H-creative, and more than P-creative.

8.3 Variations in the novelty decision method

Connecting Novelty to Expectations:

Grace et al. [2015], while evaluating surprise (expectation violations) [Brown 2012], propose four types of expectations. These should correspond to types of novelty. *Categorical:* categories are descriptions (e.g., clusters) progressively formed from experience with past designs. If a new design can't be seen as a member of any current category then it may be seen as novel. *Trend:* trend descriptions record the changes in designs over time. For example, a major change in the trend of changes for an attribute (e.g., artifact size) may signal novelty. *Relational:* with experience, correlations can be formed between attributes of a design [Chabot & Brown 1994]. These act as expected relationships between the values of attributes. Any expectation violations may signal potential novelty. *Comprehensiveness:* the observer expects that their domain knowledge (categories/clusters) is sufficient to describe and categorize new designs: i.e., it is "comprehensive and stable". However, if reclassification is required this is an indication of novelty: the amount needed might be a measure.

Using Frequency:

Frequencies might be used to represent the comparison set of designs. Of course, if a design has a frequency at all, it is not new. A set of designs can be represented by the frequency with which particular features/attributes can be found in that set. Each design to be evaluated for novelty has a set of features. A frequency can be retrieved for each of those design features from the comparison set. Those might be combined to provide a view of how novel the design is relative to the comparison set. However, this is still an approximate, and *context free*, view, as there may be no designs in the set with both feature X and feature Y, even though they are both quite frequent in the set. To refine this approach, one would need to switch to one that is *context sensitive*, similar to the n-gram methods for language recognition for text [Russell & Norvig 2009].

Sluis-Thiescheffer et al. [2016] have concerns about Shah's frequency-based novelty metric: that it is relative to a specific subset of designs; it is for the conceptual phase; and that large datasets decrease the difference in novelty between design solutions. They propose a novelty threshold that adjusts to the number of solutions for each attribute. They rank the solutions by frequency, and take only the lowest 25% of the distribution as novel: novel =1 and 0 otherwise, for each attribute. The novelty score for a design is the sum of the scores across all attributes. The threshold may need adjusting in some situations.

Research on estimating the novelty of patents [Kim et al. 2016; Uzzi et al. 2013] by using their references or technology classification codes as features, compares how frequently pairs of patent's features were assigned to other patents in history to indicate its novelty, or how the actual frequencies vary from random.

Comparison with individual designs vs. representative of set:

If designs are clustered to reflect similarity then the clusters can be used to determine relevance of a set of designs for comparison, or some 'representative of' that cluster can be formed and used for comparison.

Maher & Fisher [2012] use a "description space" using attribute-value pairs. They propose using the Euclidean *distance* between two designs: the square root of the sum of the squares of differences, assuming that differences can be made numeric. This spatial model allows finding the nearest neighbor (NN) design to a new design d, for example. They also suggest finding the average nearest neighbor distance for all the designs in the set, and then a comparison between that average and d's NN distance will reveal comparative novelty.

By clustering designs, using the K-means algorithm, and by using the centroids of each cluster (the 'average' design in the cluster), a set of representative descriptions become available. A new design can be compared to every centroid and distances calculated. These can be used to indicate novelty. As noted above, a new design might be so novel that re-clustering is required. The amount of change could be used to indicate novelty.

There are a variety of names for, and method for, obtaining descriptions that represent sets: e.g., Prototype; Centroid; Center of Gravity; Exemplars. Exemplars are selected actual instances from the set, while a Prototype is an abstracted representative. The Center of Gravity reflects the distribution of instances in the space. Typicality is associated with exemplars that share the most characteristics with other exemplars.

9. Cognitive Science Views of Similarity

Clearly the assessment of *similarity* is a key to all indicators of novelty: the intuition being that more similarity indicates less novelty. Cognitive Science has a variety of theories about similarity [Decock & Douven 2010; Hahn 2014]. Similarity reflects properties of the objects being considered; it is a relationship between two objects under a given description; it can be context dependent; it depends on how objects are represented by the agent who is judging; and the representation can depend on the current goals. In addition, the items being compared matter, as some features may become more important. Note that we see little of these issues reflected in existing design novelty measures. There are four main Cognitive Science models of similarity.

1. *Spatial* – using continuous valued dimensions.
2. *Featural* – using binary features.
3. *Structural Alignment* – using graphs or multi-place predicates.
4. *Transformational* – using transformations, with structured representations.

While 1 & 2 are not able to conveniently handle relationships (e.g., to-the-left-of) during similarity assessment, 3 & 4 can. For a comprehensive general view of similarity measures see Bergmann [2002, chapter 4].

9.1 The Spatial Model

In the commonly used Spatial model (or Geometric model), similarity is defined in terms of *distance* measured using a distance function (metric) over coordinate values on multiple continuous dimensions. Distant items should then be psychologically different (i.e., dissimilar). Dimensionality reduction can be used to provide fewer dimensions while preserving the similarity relationships. Strictly, this approach actually measures the *degree of dissimilarity* of objects in the space. The *relative similarity* can be indicated by comparing distances; *absolute similarity* can be indicated by using a predefined fixed *threshold* value.

The model not only needs selection of relevant properties to act as dimensions but also their weighting in a context dependent way. A distance function is also required, and possibly also a threshold value.

The Spatial model often does a great job of matching behavioral data, but human behavior violates the underlying axiom that the distance from A to B to be the same as from B to A. Note too that people may judge typical members of a category to be less similar to atypical members, than if the comparison is done the other way around.

The refined geometric model [Gardenfors 2004] is a modification of the spatial model, using multiple metric similarity spaces (e.g., color space; 3D Euclidean space; shape spaces; temporal space), and is known as *conceptual spaces*. The assessment context activates a subset of those spaces. This model addresses objections to the basic spatial model.

9.2 The Featural Model

This model assumes that objects are represented by sets of binary features (e.g., small; square; red). The set of features used is normally a subset of all the object's properties. The features selected will be dependent on context and purposes, and can change between comparisons.

Tversky's "Contrast Model" [1977] makes the important observation that the similarity of two objects is influenced by their differences. The three items involved are the *common* features between the two objects under consideration, as well as the two sets of features that are *distinctive* to each object: features "in A but not in B", and "in B but not in A". That is:

$$s(a, b) = \theta f(A \cap B) - \alpha f(A \setminus B) - \beta f(B \setminus A)$$

where θ, α, β act as weights, allowing similarity or dissimilarity judgments. A *salience* function f allows some features to contribute more to the comparison. Distinctive features reduce the amount of similarity established by the common features ($A \cap B$). The s(a, b) is called the Similarity Scale, with a *family* of scales depending on f and the weights. By using weights it is possible to reflect asymmetries: i.e., $s(a, b) \neq s(b, a)$.

For directional similarity comparison the model's weights can be set to emphasize the distinctive features of the New (N) item as opposed to the Existing (E) item (i.e., the asymmetry). However, it is reported that the asymmetry (the average difference in ratings) is actually only at about 5% with human subjects. Hence some researchers ignore this issue.

An alternative formulation to the one above is the “Ratio Model” where:

$$s(a, b) = f(A \cap B) \ / \ [\ f(A \cap B) + \alpha f(A \setminus B) + \beta f(B \setminus A)\ ]$$

where this can be viewed approximately as the common features divided by the total number of relevant features.

A problem with the featural model is that complexity plays a role in the comparison: note that if the features in one object are many more than the features in the other, due to complexity, then the number of features in (N – E) or (E – N) will be large and will reduce the similarity. The featural model requires a shared vocabulary of relevant features in which N and E are represented, else a comparison is meaningless.

Gomes et al. [1998] use a novelty measure with *only* distinctive features, as they are concerned with similarity as the smallest amount of difference between the design and the comparison set. Gati &Tversky [1984] showed that a few distinctive features ‘stand out’ on the background of many common features, and a few common features stand out on the background of many distinctive features.

9.3 The Structural Alignment Model

Structured representations include complex representations of objects, their parts and properties, and the *interrelationships* between them, not just lists of features or points in space. First order logic (using predicates to represent relations) or graph structures can be used to provide structured representations [Gentner & Markman 1997]. One would expect adequate representations of designs to require such representations.

Determining differences requires finding the commonalities. Structural alignment requires some matching relationships between two artifact descriptions, establishing some semantic similarity. It tries to build a “maximally structurally consistent” match across the two items, with relations and arguments in correspondence, and one-to-one mappings between elements. Matches “in place” (i.e., with matching structure) have more impact on similarity then more random matches (“out of place”).

As an example, both a car and a motorcycle have wheels: a commonality. But the difference is that cars have four wheels, and motorcycles only two. Due to the commonality, this is an “alignable” difference. However, cars have seatbelts and motorcycles do not: a “non-alignable” difference.

9.4 The Transformational Model

In this model, similarity is dependent on the amount of "effort" needed to "transform" one structured representation into the other. Transformations might include feature changes and changes of value along a dimension. This model does require commitment to a particular set of transformations. Simple, relevant, transformations are preferable: defining them is nontrivial. They can then be combined to form more complex transformations.

The transformational model can handle asymmetries, and predict human response times. It has problems as a computational model, because there's no simple way to identify the set of relevant transformations, and no simple way to decide which transformations to apply, or in which order to apply them.

9.5 For All Cognitive Science Models

All these models require decisions about some potentially psychologically relevant details: e.g., dimensions, features, weights, mapping rules for alignment, or choice of transformations. Each has strengths, and conditions for application [Hahn 2014]. From a computational point of view, spatial models connect with clustering/classification and the use of exemplars/prototypes; structural alignment models are widely used in analogical reasoning; and sets of features are a common representational method.

10. The Assumptions

It appears that there may be many assumptions being made in creativity research about novelty evaluation. The first problem is that while there is much mention of novelty in association with creativity [Sarkar & Chakrabarti 2008], the field has yet to agree on a *precise* definition, or even whether a *single* definition will suffice. There may well be an assumption that this has already been 'solved'.

There are many examples in the literature of the use of novelty scores in the evaluation of how creative a designed artifact is, and, by extension, how creative a designer is. These estimates are mostly done by human experimenters/experts: mostly using variations on the same method [Shah et al. 2003]. The assumptions appear to be that: this method is all there is; it is accurate enough; there is no need to examine this issue further.

There are few computational methods to estimate novelty. The utility of Shah's method for computational use has already been questioned [Brown 2014b]. The apparent lack of use of significantly different additional methods suggests that there is an assumption that what we already have is sufficient. Although choice of method might be due to convenience.

Another assumption is that only a single scoring method is needed. The possibility of using more approximate estimates *while designing* needs to be investigated. One can imagine intermediate design decisions being guided by one scoring method, while a more realistic method is used to select between complete designs.

In addition, it has been suggested that there are different types of novelty [Pease et al. 2001]. It may be that lack of investigation of these types reveals another assumption.

One source of information about novelty that has not been well examined is Psychology: in particular the literature about similarity detection is particularly relevant [Hahn 2014]. This may also reveal an assumption — perhaps inherited from modern Artificial Intelligence research — about the relative lack of importance of cognitively inspired approaches.

The final assumption is that existing novelty scoring methods *predict* human novelty estimates sufficiently well [Miller et al. 2017]. As novelty, and how creative an artifact is, will ultimately be judged by a human (often a user) there needs to be agreement. The work in this area has also been limited.

11. Conclusion

This paper was motivated by the following problems: we need more precise definitions of novelty; more methods for estimating novelty need to be used; there are very few computational methods; approximate novelty estimates need to be addressed; there is little discussion of types of novelty; more consideration of psychological theories is needed; there is little experimental examination of the validity of novelty evaluation methods.

We recommend that future research address the questions of: which method is the 'best' for a particular experiment or computational system; which might be the best hypothesis about what human designers do when evaluating novelty; which method most closely predicts the evaluations of human users; and whether methods might be combined. The author believes that accurate novelty assessment requires accurate descriptions, such as those reflecting structure, behavior and function in some detail. However, as I hope we have demonstrated, novelty is a complicated concept, worthy of a lot of additional study.

12. References

Bergmann R (2002) Experience Management: Foundations, Development Methodology, and Internet-Based Applications. LNAI 2432, Springer.

Besemer SP (2006) Creating products in the age of design. How to improve your new product ideas! New Forums Press, Inc.

Boden MA (1994) What is Creativity? Dimensions of Creativity. Boden MA (Ed.), The MIT Press, pp. 75-117.

Brown DC (2012) Creativity, Surprise & Design: An Introduction and Investigation. Proc. 2nd Int. Conf. on Design Creativity (ICDC2012).

Brown DC (2014a) Computational Design Creativity Evaluation, 6th Int. Conf. on Design Computing and Cognition: DCC'14.

Brown DC (2014b) Problems with the Calculation of Novelty Metrics, Workshop on Design Creativity, DCC’14.

Chabot R, Brown DC (1994) Knowledge compilation using constraint inheritance, AIEDAM 8(2), Cambridge UP, pp. 125-142.

Chakrabarti A, Khadilkar P (2003) A measure for assessing product novelty, Proc. ICED, Stockholm.

Colton S, Pease A, Corneli J, Cook M, Hepworth R, Ventura D (2015) Stakeholder groups in computational creativity research and practice, In: Besold, Schorlemmer, Smaill (Eds.), Computational creativity research: towards creative machines, Atlantis Press, pp. 3-36.

Dean DL, Hender JM, Rodgers TL, Santanen EL (2006) Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation, Journal of the Association for Information Systems 7(10), pp. 646-699.

Decock L, Douven I (2010) Similarity After Goodman, Review of Philosophy and Psychology.

Erden MS, Komoto H, van Beek TJ, D’Amelio V, Echavarria E, Tomiyama T (2008) A Review of Function Modeling: Approaches and Applications. Goel, Davis, Gero (Eds.), Special issue on Multi-modal Design, AIEDAM 22(2).

Gardenfors P (2004) Conceptual Spaces as a Framework for Knowledge Representation, Mind and Matter 2(2):9-27.

Gati I, Tversky A (1984) Weighting common and distinctive features in perceptual and conceptual judgments. Cognitive Psychology 16:341-370.

Gentner D, Markman AB (1997) Structure Mapping in Analogy and Similarity. American Psychologist 52:45-56.

Gomes P, Bento CL, Gago P (1998) A Function for Novelty Evaluation in Design, 4th Int. Conf. on Computation Models of Creative Design (HI98), JS Gero & ML Maher (Eds.), Faculty of Architecture, U. Sydney.

Grace K, Maher ML, Fisher D, Brady K (2015) Modeling Expectation for Evaluating Surprise in Design Creativity, Design Computing and Cognition '14, Springer, pp. 189-206.

Grace K, Maher, ML (2014) What to expect when you're expecting: the role of unexpectedness in computationally evaluating creativity, Proc. 4th Int. Conf. on Computational Creativity.

Hahn U (2014) Similarity. Wiley Interdisciplinary Reviews: Cognitive Science 5(3):271-280.

Jordanous A (2016) Four PPPPerspectives on computational creativity in theory and in practice, Connection Science 28(2), pp. 194-216.

Kim D, Cerigo DB, Jeong H, Youn H (2016) Technological novelty profile and invention's future impact. EPJ Data Science, 5(8).

Lopez-Mesa B, Vidal R (2006) Novelty Metrics in Engineering Design Experiments. Int. Design Conf., DESIGN 2006, pp. 557-564.

Maher ML, Fisher DH (2012) Using AI to Evaluate Creative Designs. Proc. 2nd Int. Conf. on Design Creativity (ICDC2012), pp. 45-54.

Miller S, Hunter S, Starkey E, McKay A (2017) Comparing the validity and utility of creativity metrics in engineering design research. ASME IDETC IDETC2017-68567.

Murakami T, Koyanagi T (2017) Proposal of Delta Design Map Based on Function, Behavior, Structure and User Experience for Design Ideation Support. ASME IDETC IDETC2017-68058.

Pease A, Winterstein D, Colton S (2001) Evaluating Machine Creativity. Proc. ICCBR'01 Workshop on Creative Systems, Canada, pp. 129-137.

Peinado F, Francisco Hervás R, Gervás P (2010) Assessing the novelty of computer-generated narratives using empirical metrics, Minds and Machines 20(4), pp. 565-588.

Peinado F, Gervás P (2006) Evaluation of Automatic Generation of Basic Stories. New Generation Computing 24(3), pp. 289-302.

Pérez y Pérez R, Sharples M (2001) MEXICA: a computer model of a cognitive account of creative writing. Journal of Experimental and Theoretical Artificial Intelligence 13(2), pp. 119-139.

Pérez y Pérez R, Sharples M (2004) Three Computer-based Models of Storytelling, Knowledge-Based Systems 17, pp. 15-29.

Pérez y Pérez R, Ortiz O, Luna WA, Negrete S, Peñaloza E, Castellanos V, Ávila R (2011) A System for Evaluating Novelty in Computer Generated Narratives. Proc. of 2nd Int. Conf. on Computational Creativity, pp. 63-68.

Ritchie G (2007) Some Empirical Criteria for Attributing Creativity to a Computer Program. Minds & Machines 17, pp. 67-99.

Russell S, Norvig P (2009) Artificial Intelligence: A Modern Approach, Pearson.

Sarkar P, Chakrabarti A (2008) Studying Engineering Design Creativity: Developing a Common Definition and Associated Measures, Proc. of the NSF Int. Workshop on Studying Design Creativity'08, University of Provence.

Sarkar P, Chakrabarti A (2011) Assessing design creativity. Design Studies 32, pp. 348-383.

Shah JJ, Vargas Hernandez N, Smith SM (2003) Metrics for measuring ideation effectiveness. Design Studies 24, pp. 111-134.

Simonton DK (2012) Taking the U.S. Patent Office Criteria Seriously: A Quantitative Three-Criterion Creativity Definition and Its Implications. Creativity Research Journal 24(2-3), pp. 97-106.

Sluis-Thiescheffer W, Bekker T, Eggen B, Vermeeren A, De Ridder H (2016) Measuring and comparing novelty for design solutions generated by young children through different design methods. Design Studies 43:48-73.

Srinivasan V, Chakrabarti A (2010) Investigating novelty-outcome relationships in engineering design. AIEDAM 24(2), pp. 161-178.

Tversky A (1977) Features of similarity, Psychological Review 84(4):327-352.

Uzzi, B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. Science 342(6157):468-472.

13. Selected Additional References (2020)

Ahmed F, Ramachandran SK, Fuge M, Hunter S, Miller S (2018) Interpreting Idea Maps: Pairwise Comparisons Reveal What Makes Ideas Novel, ASME. Jnl. Mech. Design, February 2019, 141(2): 021102. https://doi.org/10.1115/1.4041856

Barron F (1955) The disposition toward originality, The Journal of Abnormal and Social Psychology, 51(3), pp. 478-485. https://doi.org/10.1037/h0048073

Barto A, Mirolli M, Baldassarre G (2013) Novelty or surprise? Frontiers in psychology 4, 907. https://doi.org/10.3389/fpsyg.2013.00907

Becattini N, Borgianni Y, Cascini G, Rotini F (2017) Surprise and design creativity: investigating the drivers of unexpectedness, Int. Jnl. of Design Creativity and Innovation 5(1-2), pp. 29-47, DOI: 10.1080/21650349.2015.1090913

Bossomaier T, Harré M, Knittel A, Snyder A (2009) A semantic network approach to the Creativity Quotient (CQ), Creativity Research Journal 21(1), pp. 64-71. https://doi.org/10.1080/10400410802633517

Chulvi V, Mulet E, Chakrabarti A, López-Mesa B, González-Cruz C (2012) Comparison of the degree of creativity in the design outcomes using different design methods, Journal of Engineering Design 23(4), pp. 241-269. DOI: 10.1080/09544828.2011.624501

Coates D (2003) Watches tell more than time: product design, information and the quest for elegance, p.137 & Chapter 11, McGraw-Hill.

Dumas D, Dunbar KN (2014) Understanding Fluency and Originality, Thinking Skills and Creativity 14, pp. 56-67.

Fiorineschi L, Frillici F, Rotini F (2018) A-Posteriori Novelty Assessments for Sequential Design Sessions, Proceedings of the DESIGN 2018 15th International Design Conference, pp. 1079-1090. https://doi.org/10.21278/idc.2018.0119

Fiorineschi L, Frillici F, Rotini F (2019a) Three-Dimensional Approach for Assessing Uncommonness of Ideas. Proceedings of the Design Society: International Conference on Engineering Design 1(1), pp. 229-238. doi:10.1017/dsi.2019.26

Fiorineschi L, Frillici F, Rotini F (2019b) Orienting through the Variants of the Shah's A-Posteriori Novelty Metric. Proceedings of International Conference on Engineering Design 1(1), pp. 2317-2326. doi:10.1017/dsi.2019.238

Fiorineschi L, Frillici FS, Rotini F (2020a) Subjectivity of novelty metrics based on idea decomposition, Int. Jnl. of Design Creativity and Innovation 8(4), pp. 223-239. DOI: 10.1080/21650349.2020.1811775

Fiorineschi L, Frillici FS, Rotini F (2020b) Impact of missing attributes on the novelty metric of Shah et al., Res. Eng. Design 31, pp. 221–234. https://doi.org/10.1007/s00163-020-00332-x

Fu K, Chan J, Cagan J, Kotovsky K, Schunn C, Wood K (2013) The meaning of "near" and "far": the impact of structuring design databases and the effect of distance of analogy on design output, Journal of Mechanical Design 135(2).

Guo J, Ge Y, Pang W (2019) The underlying cognitive mechanisms of the rater effect in creativity assessment: The mediating role of perceived semantic distance, Thinking Skills and Creativity 33, 100572. doi.org/10.1016/j.tsc.2019.100572

Haller CS, Courvoisier DS, Cropley DH (2011) Perhaps There Is Accounting for Taste: Evaluating the Creativity of Products, Creativity Research Journal 23(2), pp. 99-109. DOI: 10.1080/10400419.2011.571182

Hao J, Zhao Q, Yan Y (2017) A function-based computational method for design concept evaluation, Advanced Engineering Informatics 32, pp. 237-247. https://doi.org/10.1016/j.aei.2017.03.002

Hay L, Duffy A, Grealy M (2019) The novelty perspectives framework: a new conceptualisation of novelty for cognitive design studies, Proc. of the Design Society: Int. Conf. on Engineering Design 1(1).

He Y, Luo J (2017) The novelty 'sweet spot' of invention. Design Science 3, E21. doi:10.1017/dsj.2017.23

Hiruta A (2013) Examination guidelines for patentability – Novelty and inventive step, Japan Patent Office, Examination Standards Office.

Hu J, Ma J, Feng J, Peng Y (2017) Research on new creative conceptual design system using adapted case-based reasoning technique, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 31(1), pp. 16-29. doi:10.1017/S0890060416000159

Hwang D, Wood K (2020) Assessing the Novelty of Design Outcomes: Using a Perceptual Kernel in a Crowd-Sourced Setting, Int. Conf. on Design Computing and Cognition, pp. 423-422

Jagtap S (2018) Design creativity: refined method for novelty assessment, International Journal of Design Creativity and Innovation 7(1-2), pp. 99-115. DOI: 10.1080/21650349.2018.1463176

Jiang P, Atherton M, Sorce S, Harrison D, Malizia A (2018) Design for invention: a framework for identifying emerging design–prior art conflict, Journal of Engineering Design 29(10), pp. 596-615. DOI: 10.1080/09544828.2018.1520204

Kershaw TC, Bhowmick S, Seepersad CC, Hölttä-Otto K (2019) A Decision Tree Based Methodology for Evaluating Creativity in Engineering Design, Frontiers in Psychology 10. https://www.frontiersin.org/article/10.3389/fpsyg.2019.00032

Kim D, Cerigo DB, Jeong H, Youn H (2016) Technological novelty profile and invention's future impact, EPJ Data Science 5(8). doi.org/10.1140/epjds/s13688-016-0069-1

Koo B, Shin B (2018) Applying novelty detection to identify model element to IFC class misclassifications on architectural and infrastructure Building Information Models, Journal of Computational Design and Engineering 5(4), pp. 391-400. https://doi.org/10.1016/j.jcde.2018.03.002

Kroll E, Koskela L (2017) Studying design abduction in the context of novelty, DS 87-7 Proc. of the 21st Int. Conf. on Engineering Design (ICED) 7, pp. 61-70.

Kyriakou H, Nickerson JV (2015) Novelty in Collective Design, Pre-ICIS SIGPRAG Workshop "Practice-based Design and Innovation of Digital Artifacts".

Lamb C, Brown DG, Clarke CLA (2018) Evaluating computational creativity: An interdisciplinary tutorial, ACM Computing Surveys 51(2):28.

Lubart TI (2001) Models of the Creative Process: Past, Present and Future, Creativity Research Journal 13:3-4, pp. 295-308. DOI: 10.1207/S15326934CRJ1334_07

Majid SNA, Kassim H, Razak MA (2015) Evaluating the Creativity of a Product Using Creativity Measurement Tool (CMET), Proc. Int. Conf. on Social Science Research, Malaysia. pp. 257-263.

McTeague C, Duffy A, Hay L, Vuletic T, Campbell G, Choo PL, Grealy M (2018) Insights into design concept similarity judgements, 15th Int. Design Conf. 5, pp. 2087-2098.

Ranjan B, Siddharth L, Chakrabarti A (2018) A systematic approach to assessing novelty, requirement satisfaction, and creativity. Artificial Intelligence for Engineering Design, Analysis and Manufacturing 32(4), pp. 390-414. doi:10.1017/S0890060418000148

Runco MA, Jaeger GJ (2012) The Standard Definition of Creativity, Creativity Research Journal 24(1), pp. 92-96. DOI: 10.1080/10400419.2012.650092

Sääksjärvi M, Gonçalves M (2018) Creativity and meaning: Including meaning as a component of creative solutions, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 32(4), pp. 365-379. doi:10.1017/S0890060418000112

Siddharth L, Madhusudanan N, Chakrabarti A (2020) Toward Automatically Assessing the Novelty of Engineering Design Solutions. ASME Journal of Computing and Information Science in Engineering 20(1), 10.1115/1.4044318.

Sluis-Thiescheffer W, Bekker T, Eggen B, Vermeeren A, De Ridder H (2016) Measuring and comparing novelty for design solutions generated by young children through different design methods, Design Studies 43, pp. 48-73, ISSN 0142-694X. https://doi.org/10.1016/j.destud.2016.01.001.

Stein MI (1953) Creativity and Culture, The Journal of Psychology 36(2), pp. 311-322. DOI: 10.1080/00223980.1953.9712897

Toivonen H, Gross O (2015) Data mining and machine learning in computational creativity. WIREs Data Mining Knowl. Discov. 5(6), pp. 265-275. https://doi.org/10.1002/widm.1170

Yuan X, Lee J-H (2014) A quantitative approach for assessment of creativity in product design, Advanced Engineering Informatics, 28, pp. 528-541.