ORIENTING THROUGH THE VARIANTS OF THE SHAH'S A-POSTERIORI NOVELTY METRIC

Fiorineschi, Lorenzo; Frillici, Francesco Saverio; Rotini, Federico

University of Florence

ABSTRACT

Different variants of a-posteriori novelty metrics can be found in the literature. Indeed, such a kind of assessment procedures is often used to extract useful information about creativity and/or idea generation effectiveness. In particular, the metric proposed by Shah et al. in 2003, is one of the most used and discussed in the literature. However, scholars highlighted some flaws for this metric, and some variants have been proposed to overcome them. This paper argues about the variants proposed for the a-posteriori metric of Shah et al., and proposes a selection framework to support researchers in selecting the most suited for their experimental needs. The proposed selection framework also highlights important research hints, which could pave the way for future activities. More specifically, it is still necessary to support the identification of the best-suited abstraction framework to assign weights to attributes, and the assignment of weights should be better supported as well. Moreover, this paper highlights the presence of "uncommonness of key attributes", which needs to be investigated for experimental cases where ideas missing some key attributes are present.

Keywords: Creativity, Novelty Metrics, Evaluation, Design theory, Nevelty Assessment

Contact:
Frillici, Francesco Saverio
University of Florence
Department of Industrial Design
Italy
francescosaverio.frillici@unifi.it

Cite this article: Fiorineschi, L., Frillici, F.S., Rotini, F. (2019) 'Orienting through the Variants of the Shah's A-Posteriori Novelty Metric', in *Proceedings of the 22nd International Conference on Engineering Design (ICED19),* Delft, The Netherlands, 5-8 August 2019. DOI:10.1017/dsi.2019.238

1 INTRODUCTION

The concept of novelty is one of the most important parameters when talking about innovation, creativity or "ideation effectiveness". However, it can assume different meanings, depending on the parameters on which the assessment is performed, and more importantly, depending on what is considered as the reference to establish what is novel or not novel. In particular, Shah, Vargas-Hernandez and Smith (2003) identified two distinct approaches, i.e. the "a-priori" and the "a-posteriori" ones. In the first case, it is necessary to identify a reference solution (or a reference set of solutions) to discover whether the examined ideas are more or less novel. Differently, in the second case the set of ideas to be assessed constitutes the reference itself, to count the occurrences of similar ideas generated in the same design or idea generation session. In this way, novelty is intended as a measure of the unusualness or "uncommonness" of a specific idea in relation to the same group ideas (e.g. from the same experiment). Although uncommonness is a concept of novelty which is relative to a specific set, it can be extremely useful for experimental purposes, since it allows to assess the effect of treatments (e.g. stimuli, incubation, methods, etc.), and to avoid the identification of an external reference. Indeed, the latter is one of the most critical issues of a-priori approaches: what is novel for a person may be not novel for another one. It implies to carefully select the reference by comprehensive investigations about what is the background of the considered designers, students or whatever is considered as the sample. Maybe for this reason, a-posteriori novelty metrics are often used in design research (e.g. Jansson and Smith (1991), Linsey *et al*. (2011); Vargas-Hernandez, Schmidt and Okudan (2013); Fiorineschi *et al*. (2018); Viswanathan and Linsey (2018)).
The a-posteriori novelty metric of Shah, Vargas-Hernandez and Smith (2003) (hereinafter called SNM), is one of the most largely used and acknowledged among design researchers, as confirmed by the high number of citations on Scopus. However, besides the relativeness of the concept of uncommonness, scholars recently highlighted some additional flaws of SNM, and proposed some variants with the aim to overcome them (see Section 2). Unfortunately, very often these variants have been proposed to be applied on specific experimental cases, not providing generally valid validations. Therefore, it can be difficult for researchers to orient through the different variants proposed for SNM.
The aim of this paper is to review the variants acknowledged for SNM and to highlight their advantages and criticalities. The outcomes of this work provide a first set of indications to support researchers in the identification of the SNM variant that can be used for their experimental purposes.
Section 2 shows a literature review of the currently available SNM variants. Section 3 comprehensively introduces the criticalities observed for some of the reviewed contributions, which led to the preliminary selection guideline shown in Section 4. Important considerations and possible research hints are reported in Section 5.

2 LITERATURE REVIEW

According to what is shortly reported in Introduction, Shah, Vargas-Hernandez and Smith (2003) considered two different procedures for novelty assessment, i.e. the "a-priori" and the "a-posteriori" ones. For the first approach, it is necessary to identify reference ideas for each function or attribute of the designed system, in order to assign a specific novelty score to each examined concept. Other a-priori approaches for assessing product novelty are present in literature, like that of Sarkar and Chakrabarti (2011) or the variant proposed by Jagtap (2016), where a specific framework is used to assess the degree of novelty with respect to existing systems.
Differently, SNM suggests to identify some recurring "key-attributes" among the same set of ideas which has to be examined, and to find the solutions adopted to implement them. Moreover, a normalized weight is assigned to each attribute, indicating its relative importance. Then, novelty is assessed by scoring the ideas generated for each key attribute, and scores are summed together by multiplying each of them for the related weight. More precisely the novelty score S for each attribute is calculated by (1):

$$S_{ij} = \frac{T_{ij} - C_{ij}}{T_{ij}} \times 10 \qquad (1)$$

Where T_{ij} is the total number of solutions (or ideas) conceived for the key attribute i, and design stage j; C_{ij} is the count of the current solution for the attribute i, and design stage j. Then, the overall novelty of each idea M is calculated through (2):

$$M_{SNM} = \sum_{i=1}^{m} f_i \sum_{j=1}^{n} S_{ij} p_j \qquad (2)$$

Where f_i is the weight of the attribute i, m is the number of attributes, n is the number of design stages and p_j is the weight assigned to the design stage j.
In the novelty assessment example reported in Shah, Vargas-Hernandez and Smith (2003), no particular problems or difficulties are mentioned, but some limitations have been highlighted in literature, leading some scholars to propose specific variants of the original metric (see the following subsections).
Like any a-posteriori novelty metric, SNM cannot be used to compare ideas with those of past idea generation sessions or with marketed products (Srivathsavai *et al*., 2010). According to the definitions provided by Boden (2004), SNM cannot assess historical novelty, but implements the concept of psychological novelty (i.e. what is novel for the individual mind that generated the idea) and also expands it to groups composed by multiple designers. Accordingly, SNM novelty score is dependent on the number of similar ideas conceived in a given set, therefore, the greater is the number of similar ideas, the lower is the overall novelty score (see Equation 1). Besides this and other limitations implicitly and inevitably related to a-posteriori assessment procedures based on idea infrequency, Brown (2014) highlighted other issues specifically concerning SNM. In particular he mentioned the subjectivity affecting the identification of key attributes, the subjectivity in identifying weights for each attribute, and the need to define a clear separation (in terms of design representations) between conceptual and embodiment design stages.
Here in this section, we show the literature proposals that explicitly claim to overcome some of the SNM flaws. Accordingly, the following SNM variants propose different ways to assess novelty in terms of idea uncommonness or infrequency, sharing the same fundamentals of SNM:

- The metric is applied on ideas generated during the experiment
- Attributes and functions are identified a-posteriori by one or more evaluators.

The contributions have been identified by performing a literature search on Scopus, starting from those citing Shah, Vargas-Hernandez and Smith (2003), and refining the search by keywords (novelty, assessment, etc.). Moreover, additional searches have been performed by analysing the reference lists of the reviewed contributions.

2.1 Variants of SNM still based on key attributes

2.1.1 Sluis-Thiescheffer et al.

Sluis-Thiescheffer *et al*. (2016) pointed out that when in presence of a large set of ideas to be assessed, SNM can lead to high scores, even if similar solutions appear quite often. Accordingly, they proposed to assess novelty with a binary metric (novel or not novel), where an arbitrary "expectedness" threshold is considered for the identification of less frequent (novel) solutions. The threshold selected by the cited authors is the 75th percentile for the frequencies of solutions generated for each attribute (i.e. the 25% of ideas having lower concurrencies are novel). Therefore, their metric can be represented as it follows (Equation 3):

$$M_{ST} = \begin{cases} 1 \; for \, C_i \leq P \\ 0 \; for \, C_i > P \end{cases} \tag{3}$$

Where C_i is the number of occurrences of the idea for the attribute i, and P is the percentile which can be chosen between the 0,5 and 1.

2.1.2 Fiorineschi et al.

More recently, other issues have been highlighted for SNM and its variants, which refer to the impossibility to perform correct assessments of ideas where some attributes are not implemented (Fiorineschi, Frillici and Rotini, 2018a). A proposal to deal with missing attributes has been proposed by Fiorineschi, Frillici and Rotini (2018a), where for a single design stage (the conceptual one) they propose to calculate a modified M'_{SNM} as shown in Equation 4:

$$M'_{SNM} = M_{SNM} \frac{1}{\sum_{i=1}^{p} f_i} \tag{4}$$

Where p is the number of attributes actually involved in the assessed idea, while the other terms are the same used in Equation 2. Since the numerator is the sum of the normalized weights of all attributes (i.e. 1), M'_{SNM} becomes equal to M_{SNM} when all attributes are present (i.e. when also the denominator is equal to 1).

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

2.2 Variants of SNM based on the genealogy tree

The genealogy tree (GT) is the fundamental tool proposed by Shah, Vargas-Hernandez and Smith (2003) to assess Variety. It is a hierarchical representation based on four items, i.e. physical principles (PP), working principles (WP), embodiments (EMB) and details (DET) characterizing each implemented function (Pahl *et al.*, 2007). In GT, nodes represent the number of ideas adopting a specific item variant, and lines hierarchically connect nodes belonging to the different items (see Figure 1). Moreover, each level is characterized by a different weight, which is used to take into account the different impacts that the four items are supposed to have on the Variety score.

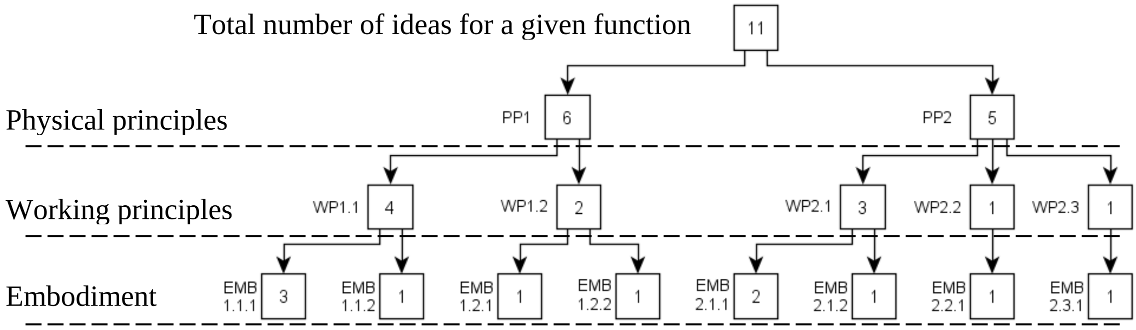

Figure 1. Example of genealogy tree (Shah et al., 2003). The "Detail" level has been removed here since not considered by the reviewed metrics.

2.2.1 Vargas-Hernandez et al.

Vargas-Hernandez, Okudan and Schmidt (2012a) pointed out that some improvements could be done in terms of capability of SNM to accurately reflect changes on the set of ideas, and to deal with boundary cases. Accordingly, they proposed two variants. The first variant proposed in Vargas-Hernandez, Okudan and Schmidt (2012a) (also used in Vargas-Hernandez, Schmidt and Okudan (2013)), focuses the attention on the WP level, and suggests a new formula for calculating novelty, as shown in Equation 5:

$$M_{VOS} = \frac{1}{C_{wp} m} \tag{5}$$

Where C_{wp} is the count of occurrences in specific WP branch (i.e. the number of nodes at the Embodiment level, which are under the same WP node), and m is the number of nodes at the WP level.
The same authors (in the same conference) presented also a very similar variant (Vargas-Hernandez, Schmidt *et al.*, 2012b), where a slightly different equation is suggested (Equation 6):

$$M_{VSO} = \frac{\mathrm{N}}{C_{wp} m} = M_{VOS} \times N \tag{6}$$

Where N is the total numbers of ideas in GT (i.e. the count of all nodes at the Embodiment level), while C_{wp} and m keep the same meaning.
Note that in both cases, instead of "Embodiment" level, the authors use the term "ideas".

2.2.2 Peeters et al.

The metric of Peeters *et al*. (2010) is substantially based on the concept that an idea can reach higher novelty scores when the difference can be observed at higher abstraction levels. Accordingly, they practically use PP, WP and EMB items in place of the SNM attributes, and uses the standardized set of weights proposed by Nelson *et al*. (2009) but continue to apply Equations 1 and 2 (Equation 7).

$$M_P = \sum_{i=1}^{m} f_i \sum_{j=1}^{o} S_{ij} w_{ij} \tag{7}$$

Where f_i is the normalized weight of function i, m is the number of functions, o is the number of items (PP, WP and STR), and w_{ij} is the normalized weight assigned to the item j, related to the function i.

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

Besides the use of function and items in place of more general key attributes, this proposal preserves the same logic of SNM. Accordingly, it presents similar issues when in presence of ideas where only a part of the required functions and items are represented/implemented.
Nevertheless, an enhanced version of M_P has been proposed (Fiorineschi, Frillici and Rotini, 2018a) to correctly deal with missing functions and/or items (Equation 8):

$$M'_P = M_P \frac{1}{\sum_{i=1}^{q} f_i \sum_{j=1}^{r} w_{ij}} \tag{8}$$

Where f_i is the normalized weight of function i, m is the number of functions, o is the number of items (PP, WP and EMB), and w_{ij} is the normalized weight assigned to the item j, related to the function i. Moreover, q is the number of functions actually involved, r is the number of items actually involved for a given function i. Similarly to what explained for Equation 4, in presence of a set of ideas implementing all functions and representing all items, M'_P and M_P lead to the same novelty scores.

2.2.3 Johnson et al.

Johnson *et al.* (2016), reported that SNM is sensitive only to differences between concepts at the embodiment level, excluding more abstract differences. However, it is important to point out that this observation is not necessarily true, since the original version of SNM referred to more general attributes, thus potentially allowing to select them at any desired abstraction level.
Johnson *et al.* (2016) also observed that M_{VOS} (Equation 3) has some limitations, i.e. the need to describe ideas at the embodiment level, same scores for siblings at the embodiment level, and possible problems with extreme scoring conditions. Accordingly, they proposed a new metric variant based on GT, where an additional level has been added, i.e. the "Strategy" one. In practice, this level is used to furtherly group solutions at an intermediate abstraction level between functions and PP. Moreover, they proposed a new formula for calculating novelty (M_J) for a single function, as reported in Equation 9:

$$M_J = \sum_{j=1}^{m} \frac{(1-P_j)s_j}{2} \tag{9}$$

Where P_j is the count of responses at the j level divided by the count of responses of its parent, i.e. the count of solutions in the node at the higher abstraction level, and s_j is the weight assigned to the level j.

3 CRITICAL FLAWS OBSERVED FOR SOME OF THE SNM VARIANTS

Here in this section, the criticalities observed in the reviewed variants are shown and explained in detail. More precisely, the following paragraphs report a detailed description of the proposals that:

- Imply additional flaws/limitations/criticalities in relation to the original version of SNM
- Failed to completely overcome the SNM issues that they were intended to face.

3.1 Variants based on key attributes

Sluis-Thiescheffer *et al.* (2016), in their experiment stressed that an idea that appears 160 times on a set of 816 ideas still get a score higher than that of an idea appearing only four times (on a set of 13 ideas). Considering the presence of 62 participants, they argued that in the first case, an idea generated 2-3 times for each participant is getting a quite higher score than an idea generated 1 time every 16 participants.
Analysing their discussion, however, it is possible to observe that they were examining two different attributes in the same design task. While one attribute is used very frequently (816 times), the other one is rarely used (only 13 times). Unfortunately, SNM doesn't take into account the uncommonness of attributes, since all attributes are expected to be implemented in the original version of Shah, Vargas-Hernandez and Smith (2003). In other words, Sluis-Thiescheffer *et al.* (2016) somehow highlighted that for certain experiments, it is useful to take into account the different occurrences of attributes (and not only of the related ideas). However, to solve this problem, it is first necessary to correctly deal with the different number of attributes composing the ideas, and secondly, it is necessary to consider the uncommonness of attributes in the final score. Concerning the uncommonness of attributes, Sluis-Thiescheffer *et al.* (2016) didn't provide any hint.
The same authors also claimed the possibility of their proposal to deal with concepts implementing only a part of the total set of key attributes, but failed in comprehensively explain how it solves the

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

problem. More precisely, for each idea, they suggest to calculate the sum of the novelties calculated for each attribute, in order to obtain the overall novelty. Nevertheless, if an idea is composed only by part of the total attributes, the minor number of scores that are going to be summed together negatively affects its overall score. This is in contrast to what recently observed for SNM, i.e. that missing attributes should affect quality score, but not novelty (Fiorineschi, Frillici and Rotini, 2018a).

However, when in presence of missing attributes the attributes' weights could be calculated on the base of their actual occurrences (the lower is the occurrence, the higher is the weight), thus reducing the subjectivity in the weight assignment. Therefore, f_i (i.e. the weight of Attribute i) in Equation 2 can be calculated with Equation 10:

$$f_i = \frac{W_i}{\sum_{i=1}^{a} W_i} \tag{10}$$

Where "a" is the number of identified attributes, and W_i is calculated by Equation 11:

$$W_i = \sum_{x=1}^{a}(C_x) - C_i \tag{11}$$

Where C_i is the number of occurrences of the attribute i, and C_x is the number of occurrences of the attribute x. Through the proposal of Fiorineschi, Frillici and Rotini (2018a), and considering Equations 10 and 11, novelty can be calculated also in presence of missing attributes, and the final score can take into account the uncommonness of attributes, but still using SNM, without the need of any drastic modification.

A further criticality that can be observed in the work of Sluis-Thiescheffer *et al.* (2016), is that the same authors also stressed that SNM could get high scores even if ideas appear frequently, and this is the main reason why they formulated their proposal (Equation 5). However, let's consider a boundary case with, for instance, a total amount of 10 ideas for an attribute where only three different types of ideas are generated, and one of these ideas appears eight times while the other two only one time each. In such a case, we easily observe that by considering the suggested 75th percentile, according to Equation 5 all the ideas are assessed as equally novel because get the same score of 1. It means that although the selection of the percentile is arbitrary, when in presence of many attributes it can be difficult to avoid unbalanced assessments among them (e.g. a certain percentile can fit well for certain attributes while not for others).

3.2 GT-based variants

3.2.1 Vargas-Hernandez et al.

The first problem that we observed in the work of Vargas-Hernandez, Okudan and Schmidt (2012a), is that they consider Equation 12 as "the original version" of the metric, where T is the total number of ideas, and C_{wp} is the number of occurrences in a WP branch:

$$S_{ij} = \frac{\mathrm{T} - C_{wp}}{C_{wp}} \times 10 \tag{12}$$

However, the recalled formula is inconsistent if compared with Equation 1. Therefore, also the flaws mentioned by the authors are inconsistent as well, because they seem to refer to Equation 12 instead of Equation 1.

Nevertheless, the main issues concerning M_{VOS} (Equation 3) are described here in the following:

- Ideas with missing items at the WP level cannot be assessed, because the metric does not consider upper levels.
- If a certain idea is expressed only in terms of WP, then missing the EMB description, $C_{wp} = 0$, Equation 3 leads to a mathematical non-sense (1 divided by 0).
- Siblings at the EMB level get the same score (Johnson *et al.*, 2016).
- The sum of novelties of ideas in GT is always 1, thus the mean novelty is strictly dependent on the number of ideas.

(The same issues observed for Equation 3 are valid also for Equation 4).

Concerning the last point, it is important to observe that Equation 3 has been used by the same authors to assess the mean novelty score in a specific experiment (Vargas-Hernandez *et al.*, 2013), but we were not been capable to verify the results. As far as we can understand, they calculated the average novelty for each idea, then calculated the average novelty for each student, and then the average novelty for each group.

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

Unfortunately, they didn't provide sufficient data to repeat the calculations. Moreover, in another work they used Equation 4 instead of Equation 3 on the same set of data, obtaining exactly the same results in terms of means and standard deviations (Vargas-Hernandez, Schmidt *et al*., 2012b).

3.2.2 Johnson et al.

Johnson *et al*. (2016), based their work on some criticalities that they observed for SNM and M_{VOS}. However, several doubts arise when reading the motivations behind the proposed metric:

- Authors used SNM by reducing it to the simple application of Equation 1, because their case study was constituted by only one function. Other a-posteriori versions based on ideas infrequency can be found in literature, which can be used when a more detailed decomposition of ideas is not needed or possible (Jansson and Smith, 1991; Linsey *et al*., 2011). It is not clear why they referred to SNM.
- Authors assert that SNM has been originally developed to be applied on sketches, but without providing any supporting reference. Actually, Shah, Vargas-Hernandez and Smith (2003) applied SNM on a case study where ideas were represented by physical prototypes.
- Authors also assert that SNM can be successfully applied only on ideas developed at the embodiment level. However, SNM is based on a very broad and general definition of "key attributes", which can potentially be applied at any abstraction level. Therefore, their assumption needs to be validated with more comprehensive verifications.
- The authors also observed that M_{VOS} leads to inconsistent results in boundary cases. More precisely they refer to the case where an ideas appears only one time in GT, i.e. a unique EMB below a specific WP with no other ideas. According to the authors, in this case M_{VOS} leads to the highest score, but it is not correct. Indeed, it seems that they wrongly consider m as the number of EMB below a certain WP. On the contrary, according to Vargas-Hernandez, Okudan and Schmidt (2012a), in this case the novelty score depends only on the number of different WPs (m), since C_{wp} is 1. An idea could get the score of 1 in M_{VOS} only in the extreme case where no other ideas are generated in the same experiment.
- Another criticality concerns the new additional level called "strategy". The actual need of this additional abstraction level is doubtful. Indeed, authors seem to use a non-conventional definition of PP-WP-EMB (e.g. they considered "Optimization" as a physical principle), and this can be the reason that led them to add a further (but non-comprehensively defined) level.

In addition, the proposed metric has a problem for boundary cases, which leads to inconsistent results. More precisely, suppose to deal with the case shown in Figure 2.

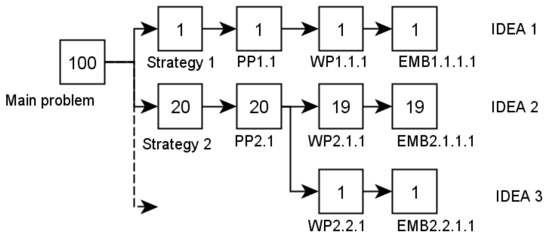

Figure 2. Generic GT after Johnson et al. (2016)

According to Equation 9 (with the s_j proposed by Johnson *et al*. (2016)), the novelty scores for the three ideas shown in Figure 2 are those listed in Table 1.

Table 1. Novelty scores calculated for ideas represented in Figure 2, by following Equation 9

	IDEA 1	IDEA 2	IDEA 3
M_J	4,95	4,08	5,42

It is possible to observe that Idea 3 gets a score higher than Idea 1. However, the latter is composed by the most uncommon solutions for each level, while Idea 3 is composed by a strategy and a PP that are used quite often. Therefore, the score is logically inconsistent.

The problem resides in the term $1-P_j$ of Equation 7. More precisely, P_j is the ratio between the occurrences at level j and the occurrences at level j-1. Therefore, for Idea 1, at PP, WP and EMB

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

levels P_j is equal to 1, thus nullifying the partial novelty scores of these levels in the summation. The same effect is present on WP and EMB levels for Idea 3, but the relative effect is smaller.

4 SUGGESTIONS FOR METRIC SELECTION

As shown by the contributions reviewed in Section 3, different experiments may have different requirements and may lead, therefore, to face some lacks of the original version of SNM. Nonetheless, SNM offers a robust base, which can be enhanced and developed to adapt it to any specific objective. More drastic modifications can also be proposed, but in this way it is possible to lose part of the robustness characterizing the original metric, and to introduce unprecedented problems. According to our critical review, this is what happened in the proposals of Sluis-Thiescheffer *et al*. (2016), Vargas-Hernandez, Okudan and Schmidt (2012a) and Johnson *et al*. (2016).

Due to the presence of different SNM alternatives, it can be difficult to orient through the available options and to discern the most robust ones according the specific objective of an experimental activity. To provide a first help in that sense, we describe here a selection algorithm that could support researchers in identifying the most suitable metric for their experimental purposes.

A first simplification comes from the criticalities highlighted in this paper for the metrics of Sluis-Thiescheffer *et al*. (2016), Vargas-Hernandez, Okudan and Schmidt (2012a) and Johnson *et al*. (2016), which lead us to exclude them. Then, a selection framework is proposed in Figure 3, where also three possible hints for future developments and research activities are highlighted.

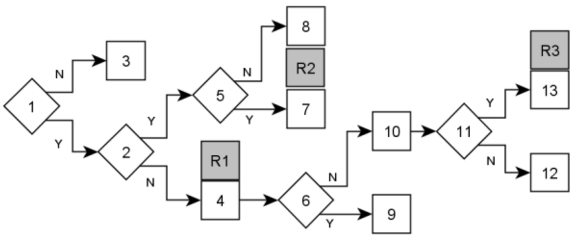

Figure 3. Framework for a-posteriori metric selection. Hints for further developments and research are represented by grey boxes (see Table 2 for steps description). Y = Yes, N = No.

Table 2. Steps of the selection framework represented in Figure 3.

N°	Description
1	Is it possible/useful to decompose/represent ideas in terms of key attributes or functions?
2	Is it possible to decompose in terms of functions PP, WP and EMB?
3	Use metrics based on infrequency of the overall ideas (e.g. Linsey *et al*. (2011)).
4	Use generic attributes.
5	All functions and/or items are always present and represented in each idea?
6	All attributes are always present in each idea?
7	Use M_P
8	Use M'_P
9	Use M_{SNM}
10	Use M'_{SNM}
11	Is it important to consider the uncommonness of attributes?
12	Assign arbitrary weights
13	Assign weights which are proportional to the attributes' occurrences (Eq. 10 and 11).
R1	Maybe other abstraction frameworks can be used in place of PP-WP-EMB.
R2	Which is the right set of weights for the abstraction levels?
R3	The proposed solution needs to be validated

The proposed framework is thought for researchers that intend to use the a-posteriori procedure proposed by Shah, Vargas-Hernandez and Smith (2003), but are not sure about the actual compatibility with their experimental set.

The first verification to be performed is about the possibility (or the actual need) to analyse ideas by decomposing them in terms of attributes or functions (Step 1 in Figure 1). If it is not possible or

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

needed, simpler metrics exist that could be used in place of SNM, which also provide agile procedures to perform the assessments (Linsey *et al*., 2011).

If the decomposition is needed/possible, it should be preferable to rely on a reference abstraction framework, in order to reduce subjectivity in the identification of attributes. Accordingly, Step 2 asks whether this particular decomposition is possible (according to the available set of ideas and the experimental requirements) or not. If this particular decomposition is possible, M_P can be used if all functions and items are always present in each idea (Steps 5 and 8), while M'_P should be used when in presence of missing functions and/or items (Steps 5 and 7). However, it is still not clear which is the right set of weights to be assigned to different functions (when multiple ones are present) and the different abstraction levels (R2 in Figure 3). If the specific decomposition mentioned in Step 2 is not possible considering the available alternatives, we suggest to identify key attributes as in the original SNM version. However, as highlighted by R1 in Figure 3, other abstraction frameworks could be investigated for being applied in SMN. Brown (2014) referred to Function-Behaviour-Structure (Gero, 1990), but also the SAPPHIRE framework of Srinivasan and Chakrabarti (2009) could provide interesting hints.

When using general key attributes, it is important to verify whether they are always present in each idea to be assessed (Step 6). If key attributes are presents, M_{SNM} can be used in its original version (Step 9). Instead, in presence of ideas with different amounts of implementing attributes, M'_{SNM} can be used (Step 10) but it is necessary to perform a further verification (Step 11). Indeed, according to the need expressed by Sluis-Thiescheffer *et al*. (2016) and discussed in Section 3.1, when in presence of variable occurrences of attributes, in certain cases it could be useful to consider their uncommonness in the final novelty score. If the uncommonness of attributes is not needed, weights of attributes can be assigned arbitrarily (Step 12). Differently, it is possible to follow Equations 10 and 11 (Step 13). However, this is only a preliminary suggestion which should be validated in experimental applications (R3).

5 DISCUSSIONS AND CONCLUSIONS

The contributions reviewed in this paper constitute the current state of the art in terms of SNM-based novelty assessments, and somehow try to enhance the original metric and/or to fill existing gaps. Unfortunately, some of these contributions presented some criticalities that avoid the possibility to use them for comprehensive novelty assessments. Nevertheless, some of the reviewed works triggered further reflections about SNM and its limits. Indeed, it has been highlighted here that in certain cases, where generated ideas implement a different number of attributes, it could be useful to consider also the uncommonness of attributes when calculating the final novelty score of each idea. Accordingly, we have proposed Equations 10 and 11, which if used concurrently with the proposal of (Fiorineschi, Frillici and Rotini (2018a), could provide a first preliminary solution to fulfil this need.

The main contribution of this paper is a thorough and critical review of the SNM variants, which led to a systematic framework aimed at supporting researchers in the selection of the most suited SNM-based metric for their experiments. Therefore, this work is intended to support the design research, and to provide useful hints for future activities about creativity and innovation assessment.

But there are several other argument about the assessment of uncommonness, which need to be investigated in depth. Indeed, as highlighted by Fiorineschi, Frillici and Rotini (2018b), when in presence of sequential design sessions, the reference universes of ideas should be selected with care, in order to correctly evaluate the effects of the investigated treatments. A particular example can be found in a work of some of the same authors, where an experiment constituted with multiple design sessions was performed (Cascini *et al*., 2018).

Moreover, even with singular sessions, when in presence of more groups of designers to be evaluated (e.g. control group and analysis group), it is not always clear which is the best way to compose the reference universes. Indeed, in some cases (e.g. in Vargas-Hernandez, Schmidt and Okudan, (2013), the reference universe is created by merging the sets generated by the different groups. However, in this way it is implicitly assumed that all groups have the same background and then could be potentially capable to produce the same variety of ideas. This assumption could be admissible or not, but should be validated for each experiment. Indeed, it is possible that while an idea is rarely generated in a group, the same idea can be used quite often in another group. Therefore, by merging the universes, the effect of the investigated treatment can be attenuated. The right strategy to follow is surely case sensitive and depends on the experimental objectives. But to provide a comprehensive

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press

guideline for a-posteriori novelty assessment, this argument and the other highlighted here should be investigated in deep.

REFERENCES

Boden, M.A. (2004), *The Creative Mind: Myths and Mechanism*, 2nd ed., Routledge, London, available at: https://doi.org/10.4324/9780203508527.

Brown, D.C. (2014), "Problems with the Calculation of Novelty Metrics", *Proceedings of the 6th Int. Conf. on Design Computing and Cognition*, available at: http://web.cs.wpi.edu/~dcb/Papers/DCC14/DCC14-Brown-Novelty-workshop.pdf.

Cascini, G., Fiorineschi, L. and Rotini, F. (2018), "Investigating on the Re-use of Conceptual Design Representations", *International Design Conference - Design 2018*, pp. 1009–1020.

Fiorineschi, L., Frillici, F.S. and Rotini, F. (2018a), "Issues Related to Missing Attributes in A-Posteriori Novelty Assessments", *International Design Conference - Design 2018*, pp. 1067–1078.

Fiorineschi, L., Frillici, F.S. and Rotini, F. (2018b), "A-Posteriori Novelty Assessments for Sequential Design Sessions", *International Design Conference - Design 2018*, pp. 1079–1090.

Fiorineschi, L., Frillici, F.S., Rotini, F. and Tomassini, M. (2018), "Exploiting TRIZ Tools for enhancing systematic conceptual design activities", *Journal of Engineering Design*, Vol. 29 No. 6, pp. 259–290.

Gero, J.S. (1990), "Design Prototypes : A Knowledge Representation Schema for Design", *AI Magazine*, Vol. 11 No. 4.

Jagtap, S. (2016), "Assessing Design Creativity: Refinements to the Novelty Assessment Method", *International Design Conference - DESIGN 2016*, pp. 1045–1054.

Jansson, D.G. and Smith, S.M. (1991), "Design fixation", *Design Studies*, Vol. 12 No. 1, pp. 3–11.

Johnson, T.A., Caldwell, B.W., Cheeley, A. and Green, M.G. (2016), "Comparison and Extension of Novelty Metrics for Problem-Solving Tasks", *Proceedings of the ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference IDETC/CIE 2016*, available at: https://doi.org/10.1115/DETC2016-60319.

Linsey, J.S., Clauss, E.F., Kurtoglu, T., Murphy, J.T., Wood, K.L. and Markman, a. B. (2011), "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods", *Journal of Mechanical Design*, Vol. 133 No. 3, p. 031008.

Nelson, B.A., Wilson, J.O., Rosen, D. and Yen, J. (2009), "Refined metrics for measuring ideation effectiveness", *Design Studies*, Elsevier Ltd, Vol. 30 No. 6, pp. 737–743.

Pahl, G., Beitz, W., Feldhusen, J. and Grote, K.H. (2007), "Engineering Design" 3rd Ed, Springer-Verlag London, available at: https://doi.org/10.1007/978-1-84628-319-2.

Peeters, J., Verhaegen, P.A., Vandevenne, D. and Duflou, J.R. (2010), "Refined Metrics for Measuring Novelty in Ideation", *Proceedings of IDMME - Virtual Concept 2010*, pp. 1–4.

Sarkar, P. and Chakrabarti, A. (2011), "Assessing design creativity", *Design Studies*, Elsevier Ltd, Vol. 32 No. 4, pp. 348–383.

Shah, J.J., Vargas-Hernandez, N. and Smith, S.M. (2003), "Metrics for measuring ideation effectiveness", *Design Studies*, Vol. 24 No. 2, pp. 111–134.

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

Srinivasan, V. and Chakrabarti, A. (2009), "SAPPHIRE-An Approach to Analysos and Synthesis", *International Conference on Engineering Design - ICED'09*, Stanford, CA, USA, pp. 417–428.

Srivathsavai, R., Genco, N., Katj and Seepersad, C.C. (2010), "Study of Existing Metrics Used in Measurement if Ideation Effectiveness", *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2010*.

Vargas-Hernandez, N., Okudan, G.E. and Schmidt, L.C. (2012a), "Effectiveness Metrics for Ideation: Merging Genealogy Trees and Improving Novelty Metric", *Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2012*, available at: https://doi.org/10.1115/DETC2012-70295.

Vargas-Hernandez, N., Schmidt, L.C. and Okudan, G.E. (2012b), "Systematic Ideation Effectiveness Study of TRIZ", *Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2012*, pp. 1–10.

Vargas-Hernandez, N., Schmidt, L.C. and Okudan, G.E. (2013), "Systematic Ideation Effectiveness Study of TRIZ", *Journal of Mechanical Design*, Vol. 135 No. 10, p. 101009.

Viswanathan, V.K. and Linsey, J.S. (2018), "Role of Sunk Cost in Engineering Idea Generation : An Experimental Investigation", *Journal of Mechanical Design*, Vol. 135 No. December 2013, pp. 121002 1–12.

https://doi.org/10.1017/dsi.2019.238 Published online by Cambridge University Press