THREE-DIMENSIONAL APPROACH FOR ASSESSING UNCOMMONNESS OF IDEAS

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

University of Florence

ABSTRACT

A posteriori novelty metrics are often used in design research, in order to extract important information about creativity. However, different assessment approaches can be found in the literature, each of them with related pros and cons. In particular, weighted uncommonness, overall uncommonness and uncommonness across groups are the three main families of a-posteriori novelty metrics identified in this paper. Each of the considered literature metrics can provide specific types of information about the uncommonness of ideas, but in certain experimental circumstances, it could be difficult to rapidly identify the best-suited approach. This paper proposes an integrated procedure where the advantages offered by the three families of metrics can be applied concurrently. A generic case study is used for a first application of the proposal, and the obtained results show that a more comprehensive set of information about a-posteriori novelty can be extracted. In particular, novelty data from the three families of metrics are extracted in a single assessment process.

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

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

1 INTRODUCTION

Creativity is one of the most recurring arguments in design research, where different parameters can play key roles. Several metrics and definitions have been proposed during years (Sarkar and Chakrabarti, 2008), where novelty constitutes one of the most important parameters (Corazza and Agnoli, 2016; Sarkar and Chakrabarti, 2011). This is the case of the metrics proposed by Shah, Vargas-Hernandez and Smith (2003), who identified two main families of approaches for novelty assessments: "a-priori" and "a-posteriori". In a-priori approaches, the novelty of ideas is evaluated by taking into account a reference solution (or a reference set of solutions) identified by the evaluators. However, what is novel for a person may be (or appear as) not novel for another person, thus leading to a certain subjectivity in the assessments. On the contrary, a-posteriori approaches aim at measuring the unusualness or "uncommonness" of ideas, within a group of ideas generated in specific sessions. Therefore, the set of ideas to be assessed constitutes the reference sample. Consequently, since a-posteriori approaches do not compare ideas with those of past idea generation sessions or with marketed products (Srivathsavai *et al.*, 2010), the obtained scores represent only a relative novelty.

Nevertheless, a-posteriori novelty metrics are very often considered by scholars in their experiments, as demonstrated by the numerous citations received by the metric proposed by Shah, Vargas-Hernandez and Smith (2003). Literature also acknowledges some variants of this metric (Fiorineschi *et al.*, 2018a), which however are based on the same fundamentals, i.e. calculating the mean infrequency of the idea variants used for the attributes in which each overall concept can be decomposed. Hereinafter, this group of metrics is called "SNM-based", where SNM stands for "Shah's Novelty Metric". Furthermore, other two a-posteriori assessment approaches emerged from our literature review, which are those introduced by Jansson and Smith (1991) and Lopez-Mesa and Vidal (2006).

In this paper, we discuss about potentialities and lacks of the three different approaches, highlighting the advantages coming from the exploitation of the related three different views of novelty. Accordingly, the objective of the paper it to present a proposal aimed at exploiting the advantages offered by the three approaches (i.e. allowing a three dimensional a-posteriori novelty assessment).

The contents of the paper are structured as follows. Section 2 reports a literature review of a-posteriori novelty metrics, showing existing variants and observed issues. Section 3 comprehensively introduces advantages and disadvantages characterizing the three approaches. In Section 4, an integration proposal is presented to provide a new perspective for assessing uncommonness of ideas. Discussions are shown in Section 5, together with the impact expected by this contribution, and the possible research hints. Eventually, conclusions are reported.

2 BACKGROUND

2.1 Assessing novelty through the overall solution infrequency

The metric used by Linsey *et al.* (2011) is based on the infrequency of overall solutions. More precisely, novelty is calculated by Equation 1.

$$Novelty = 1 - \frac{number\ of\ very\ similar\ concepts}{total\ number\ of\ concepts} \tag{1}$$

The identification of similar concepts is made by generating different "bins", each of them representing groups of similar solutions. Therefore, the number of similar concepts is the number of ideas in a certain bin, while the total number of concepts is the sum of ideas contained in all the created bins. Accordingly, the number of bins is also used to assess the variety of generated ideas.

Equation 1 constitutes the fundamental part of the originality metric used by Jansson and Smith (1991).

2.2 SNM-based family of a-posteriori novelty metrics

Similarly to the metric described in the previous section, SNM-based approaches are mainly grounded on solutions' infrequency. However, the a-posteriori version originally proposed by Shah, Vargas-Hernandez and Smith (2003) suggests to identify recurring "key-attributes" among the set of assessed ideas, and to find the solutions adopted to implement them. Moreover, a normalized weight is assigned to each attribute and to the design stages, in order to consider their relative importance.

To assess novelty, ideas are firstly scored in reference to each key attribute (Equation 2):

$$S_{ij} = \frac{T_{ij} - C_{ij}}{T_{ij}} \times 10 \qquad (2)$$

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. Afterwards, the overall novelty of each idea M is calculated through (3):

$$M_{SNM} = \sum_{i=1}^{m} f \sum_{j=1}^{n} S_{ij} p_j \qquad (3)$$

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.

The case study reported in Shah *et al.* (2003), concerns a single-stage design task where designers (students) identified 46 ideas characterized by three key attributes. In the recalled example, no particular problems or difficulties are mentioned. However, Sluis-Thiescheffer *et al.* (2016) highlighted that, in presence of large sets of ideas, 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 threshold is considered for the identification of less frequent (novel) solutions for each attribute. The observation made by Sluis-Thiescheffer *et al.* is a mere consequence of the different numbers of solutions populating each attribute in their experiment (i.e., some of the overall concepts generated by the considered sample "missed" to implement some attributes). Moreover, they failed in clearly explain how to extract overall novelty scores of ideas that are constituted by only a part of the required attributes.

A proposal to comprehensively deal with assessment tasks involving missing attributes has been proposed by Fiorineschi, Frillici and Rotini (2018b). More in particular, in presence of a single design stage (the conceptual one), they suggest to divide M_{SNM} by the sum of the normalized weights of the attributes actually involved in the concept to be assessed, and then obtaining the modified overall novelty M'_{SNM}. In such a way, when all attributes are present, M'_{SNM} and M_{SNM} lead to the same values.

Other particular variants of SNM exist, which are based on the so called "genealogy tree" (GT). GT is a fundamental tool to assess Variety according to the approach proposed by Shah, Vargas-Hernandez and Smith (2003). In GT, by taking as a reference the framework of Pahl *et al.* (2007), examined ideas are decomposed according to different abstraction levels, i.e. main functions, physical principles (PP), working principles (WP), embodiment (EMB) and structural details (DET). Subsequently, a hierarchical tree (i.e. the GT) is built, where each node indicates the number of ideas adopting a specific solution variant at a specific abstraction level or item (PP, WP, EMB, DET), and lines hierarchically connect nodes belonging to the different levels (or items). 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.

The novelty metric proposed by Peeters *et al.* (2010) (M_{GT}) is substantially based on the hypothesis that an idea has high novelty scores when differences are observed at high abstraction levels. Accordingly, they use PP, WP and EMB items instead of the SNM attributes, and the standardized set of weights proposed by Nelson *et al.* (2009), while the overall novelty is still calculated through Equations 1 and 2 (Equation 4).

$$M_{GT} = \sum_{i=1}^{m} f \sum_{j=1}^{o} S_{ij} w_{ij} \qquad (4)$$

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.

Fiorineschi, Frillici and Rotini (2018b) proposed an enhanced version even for this metric to correctly deal with missing functions and/or items. Similarly to what explained for Equation 3, they suggest to calculate a modified overall novelty M'_{GT} by considering the normalized weights of functions and items actually involved in the assessed solution. Therefore, in presence of ideas implementing all functions and representing all items, M'_{GT} and the original M_{GT} lead to the same scores.

Another GT-based variant of SNM is that of Vargas-Hernandez, Okudan and Schmidt (2012), which focuses the attention only on the WP level, and proposes a new formula for calculating S. However, Johnson *et al.* (2016) observed that the recalled metric has some limitations, i.e. the need to have ideas described at embodiment level, same scores for siblings at the embodiment level, and possible problems with extreme scoring conditions. According to the above limitations and other drawbacks acknowledged for SNM (among which, the lack of consideration for uncommonness of PPs and WPs), Johnson *et al.* (2016) proposed a new metric variant based on GT, where an additional level was 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 for a single function.

2.3 The proposal of Lopez-Mesa and Vidal.

Lopez-Mesa and Vidal (2006) claimed to use the concept of non-obviousness to evaluate the creativity styles of individuals. They analysed each generated idea in terms of Functions (F), Conceptual Structure (S) and Detailed Structure (D) to identify recurrent ideas among the sample of design teams participating to their experiment. More precisely, for each team they calculated different percentage values, according to Equation 5:

$$F_{G,i} \% = \frac{f_{G,i}}{f_{G,i} + s_{G,i} + d_{G,i}} \times 100 \tag{5}$$

Where,

- $F_{G,i}$ is the percentage of ideas of the team "G", which have been produced by teams "i" at the Function level;
- $f_{G,i}$ is the number of solutions proposed by the team "G" at the F level and that appear in different "i" teams;
- $s_{G,i}$ and $d_{G,i}$ represent the number of solutions proposed by the team "G" respectively at S and D levels, and that appear in different "i" teams.

The equation is calculated for $1 \leq i \leq$ "total number of teams" involved in the experiment.

Moreover, similarly to Equation 5, another series of percentages is calculated, i.e. the $S_{G,i}$ (for the Conceptual Structure level), where $s_{G,i}$ substitutes $f_{G,i}$ at the numerator.

Subsequently, the authors plotted each calculated percentage in two distinct graphs, respectively with $F_{G,i}$ and $S_{G,i}$ on the y-axis, and the number of teams sharing the same ideas on the x-axis.

3 PROS AND CONS OF THE CONSIDERED A-POSTERIORI APPROACHES

Besides the acknowledged limits of assessing novelty in terms of uncommonness of ideas (See Section 1), each reviewed approach has different advantages and disadvantages. A list of the most evident pros and cons of the considered metrics is shown in Table 1, and a more comprehensive description of them is reported in the following paragraphs.

Table 1. General pros and cons of the reviewed metrics

Metric/approach	Pros	Cons
Jansson and Smith	Easy to be used. Unique concepts can be highlighted.	Subjectivity in the definition of bins. Differences at very low abstraction levels may sensibly affect the scores.
SMN-based	The score takes into account the importance or the abstraction level of attributes. GT versions reduce subjectivity in the definition of attributes.	The overall score may fail in highlighting unique combinations of attributes' solutions.
Lopez-Mesa and Vidal	Comprehensive analysis of the frequency of solutions generated by teams involved in the experiment.	Novelty depends on the number of teams involved in the experiment. Misleading results for boundary cases. The two lower abstraction levels are not sufficiently defined.

3.1 Jansson and Smith

The approach proposed by Jansson and Smith (1991) is quite easy to be used, indeed it is sufficient that two or more evaluators independently perform the assessment (generating potentially different numbers

of bins) to extract a mean novelty score for each team, group or individual. As suggested by the same authors, an Inter Rater Reliability test should be performed to verify the robustness of the assessment.

According to the metric and the related assessment procedure, concepts that are considered unique (i.e. only one concept in a bin), get the highest possible score, whose value depends on the total number of generated ideas. However, depending on the identification of the bins and then on the expertise and background of the evaluators, it is possible that an idea is unique only because details from low abstraction levels are considered (e.g. a mere structural detail). This can be in contrast with literature asserting that the abstraction level should play an important role in novelty assessments: the higher is the abstraction level in which the difference is observed, the higher is the novelty (Jagtap, 2016; Sarkar and Chakrabarti, 2011).

Moreover, in presence of incomplete solutions, e.g. solutions that implement only a part of the required functions, the implications of the metric application are not clear. Indeed, if incomplete solutions are neglected in the assessment, some important data can be lost (Fiorineschi et al., 2018a). On the contrary, if incomplete solutions are not discarded, it is not clear in which bin they should be placed. If a specific bin is generated for a single incomplete concept, this one obtains the highest novelty. This outcome might be correct or not, depending on the principles by which other bins have been identified. For instance, if the designers are asked to produce ideas implementing more functions, what does happen if a particular idea implements only one function and neglects the others? It would be not possible to extrapolate information about the neglected functions. Then it would be difficult to understand whether it should be placed in an existing bin or in another bin specifically created. In the latter case, the incomplete idea gets the highest novelty score since the metric does not allow to directly discern different behaviours implementing the same function but it depends on the subjective rationale under which bin are created.

3.2 SNM-based metrics

Also novelty scores calculated by SMN-based metrics depend on the number of similar ideas conceived in a given set, therefore (except for the proposal of Sluis-Thiescheffer et al. (2016)) the greater is the number of similar ideas, the lower is the overall novelty score.

According to Brown (2014), the original version of SMN is affected by subjectivity in the identification of key attributes and weights of each attribute. During the years, several issues have been highlighted for SNM and, accordingly, some improvements and modifications have been proposed, as already shown in Section 2.2. Concerning the identification of key attributes, the metrics relying on GTs somehow reduce the subjectivity since they refer to a pre-defined abstraction framework. Nevertheless, other abstraction frameworks could be potentially used (e.g. the Function - Behavior -Structure by Gero, 1990), and the weights assigned to abstraction levels still remain arbitrary.

It is possible to assert that SMN and its developments can be considered as a more comprehensive application of the concept of ideas' infrequence (accordingly, Equations 1 and 2 are extremely similar). Moreover, Fiorineschi, Frillici and Rotini (2018a) highlighted that specific strategies should be used with SNM-based metrics in case of multiple design sessions.

However, Equation 2 implies that even unique concepts, i.e. unique combinations from the variety of ideas generated for each attribute, get a score that does not represent concept uniqueness but that is dependent on the occurrences considered for attributes' ideas and weights.

For example, considering the case study analysed in Shah, Vargas-Hernandez and Smith (2003), the set of 46 generated ideas is constituted by only 18 different combinations of attribute variants, and each combination appears with a different frequency. In other words, there are concepts whose particular combinations appear only one time (unique combinations), while there are other concepts whose combinations appear more frequently. However, unique combinations can be obtained also from frequently used attributes' ideas (e.g. Entry n°6 in the case study of Shah, Vargas-Hernandez and Smith (2003)). Consequently, in these cases the final score cannot be representative of the uniqueness of the combination, because it is affected by the frequency of the solution used for each attribute as well as by the related weights (this is why the mentioned Entry n°6 gets a quite low novelty score).

Similar observations can be made for SMN variants and derivatives introduced in Section 2.2.

3.3 The proposal of Lopez-Mesa and Vidal

According to Sarkar and Chakrabarti (2011), this assessment method needs to be validated by means of further experimental applications, also in presence of more extended samples. Indeed, the number of

times in which Equation 5 is applied, is strongly dependent on the number of teams involved in the experiment. Moreover, the difference between Conceptual Structure and Detailed Structure should be defined more comprehensively. Accordingly, when operating with outcomes from conceptual design tasks, it can be extremely difficult to discern more than one abstraction level in terms of structure.

Besides the flaws mentioned above, it seems that Equation 5 leads to misleading scores in boundary cases. For instance, let us consider the example where a team (among the others) is the sole that generated a specific overall concept not appearing in other teams at any level, and repeated it N times. It means that the same functions and the same structures are repeated N times as well. Therefore, $f_{G,1}$ $s_{G,1}$ and $d_{G,1}$ assume the same value, i.e. the total number of solutions generated by the team (N). Consequently, both $F_{G,1}$ and $S_{G,1}$ assume the same value, i.e. 33,3%, while the expected value should be 100% in this particular case.

Nevertheless, notwithstanding these issues we can assert that Lopez-Mesa and Vidal presented a different way for evaluating the attitude of teams in generating uncommon ideas. In fact, instead of comparing novelty scores from ideas generated by the teams, they compare how many teams generate the same ideas.

4 THREE-DIMENSIONAL APPROACH

It is impossible to suggest the best a-posteriori assessment approach, also because most of the contributions reviewed in Section 2 came up from specific experimental needs and therefore are best suited for specific situations. For this reason, novelty assessment tasks performed by relying on a specific approach might be sometimes misleading and/or could partially hidden some experimental outcomes. Unfortunately, the differences among the considered approaches also concern how each idea/concept is analysed. Therefore, applying the three metrics to perform concurrently three different assessments, would involve too much experimenters (especially when the sample of ideas is huge). In addition, the outcomes obtained from the three different metrics are not directly comparable because they lead to different values and involve different parameters. Nevertheless, the different views offered by the three metrics could be certainly useful to better highlight different experimental results. This section presents an integration proposal that tries to exploit the three views by reducing the assessment efforts.

4.1 Integration proposal

To exploit the advantages of the three metrics, we propose an integrated approach mainly based on the SNM-based strategy. Accordingly, we still suggest to decompose ideas in terms of attributes but also to extract three different types of information about uncommonness:

- Weighted uncommonness: it is calculated by following one of the SNM-based metrics. In particular, if selecting one of the GT versions (e.g. Peeters *et al.* (2010)), it is possible to refer to a pre-defined set of attributes, already ordered in terms of abstraction levels.
- Overall uncommonness: it can be calculated by considering the specific combinations of attributes' ideas that are used to compose each overall concept. Each particular combination automatically constitutes a virtual bin, and Equation 1 can be applied. Therefore, the occurrences of each particular combination correspond to the number of ideas in a bin. In this way, the number of bins is dependent on the attributes and on the particular set of generated ideas.
- Uncommonness across groups (or individuals): it can be calculated by counting the number of groups (or individuals) that generated similar solutions (i.e. that particular combination of solution variants for each attribute).

4.2 Generic example

When performing design experiments, it is quite frequent to consider samples composed by two or more groups of designers (e.g. Vargas-Hernandez, Shah and Smith (2010); Vargas-Hernandez, Schmidt and Okudan (2013); Tsenn *et al.* (2014); Cascini, Fiorineschi and Rotini (2018); Koh and Lessio (2018)).

Accordingly, we consider here an example that simulates a situation with two groups of designers performing the same design task. Let us suppose that the two groups come from different firms, and the starting objective of the experiment is to evaluate the differences between design outcomes in terms of creativity. For the scope of this paper, the attention is limited to novelty, intended as uncommonness of generated ideas.

As a simulation, thirty virtual ideas for each group were generated randomly and represented by a code which is made of three numbers: the first one identifies the specific PP variant, the first one together with the second one identifies the specific WP variant, and all the three numbers together identify the specific EMB variant. In this way, it is implicitly assumed that a specific WP cannot belong to different PPs, and, similarly, that a specific EMB cannot belong to different couples of PPs and WPs. Table 2 shows the set generated for a single function together with the related novelty scores for both the weighted uncommonness and the overall uncommonness, according to the proposal introduced in Section 4.1. The weighted uncommonness has been calculated with the metric represented by Equation 4, by considering each group of ideas as a distinct universe. Similarly, also the overall uncommonness has been calculated by considering groups as distinct universes. Moreover, the same number of ideas for each group was imposed.

Table 2. Randomly-generated set of ideas and related novelty scores

Code	Occurr. Group1	Occurr. Group 2	Weighted uncom. Group 1	Weighted uncom. Group 2	Overall uncom. Group 1	Overall uncom. Group 2
111	7	3	6,37	7,8	7,67	9
112	3		6,5	-	9	-
121	2	3	7,33	7,5	9,33	9
122		1	-	7,57	-	9,67
123		2	-	7,53	-	9,33
211	2	3	9,33	9	9,33	9
311	1	3	9,47	5,8	9,67	9
312		2	-	5,83	-	9,33
321	1	8	9,47	5,03	9,67	7,33
322		1	-	5,27	-	9,67
323		2	-	5,23	-	9,33
331		1	-	6,17	-	9,67
332		1	-	6,17	-	9,67
411	1		9,67	-	9,67	-
511	2		8,83	-	9,33	-
512	1		8,87	-	9,67	-
521	1		9,07	-	9,67	-
611	1		8,47	-	9,67	-
612	1		8,47	-	9,67	-
613	3		8,4	-	9	-
711	1		9,07	-	9,67	-
721	1		9,07	-	9,67	-
731	1		8,97	-	9,67	-
732	1		8,97	-	9,67	-

By observing the values listed in Table 2, it is possible to understand quite rapidly that Group 2 generated a lower variety of ideas and that a greater number of unique combinations of PP-WP-EMB appears in Group 1 ("Occur. Group 1" vs "Occur. Group 2" columns). Moreover, the reader can observe, for instance, that some ideas have been generated by both groups but with different occurrences, and that the groups used certain PPs and WPs with different frequencies. Table 2 also shows that the uncommonness across groups can provide a valid help in allowing a rapid identification of these characteristics, especially in a graphical form.

To better highlight the advantages of using the three uncommonness scores concurrently, the latter could be represented in a single graph, as shown in Figure 1, where the weighted uncommonness is on the x-axis, the overall uncommonness on the y-axis and the uncommonness across groups is represented as the diameter extension of the circles representing each specific combination of PP-WP-EMB

In particular, this specific graph presents an off-limits zone, where it is not possible to have any point. This is a direct consequence of the metrics used along the two axis, together with the logical impossibility of occurrences of an overall concept to be higher than the maximum occurrences of the related solution variants for each attribute.
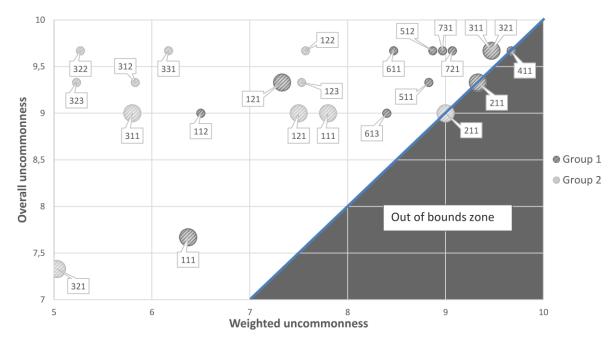
Figure 1. Concurrent representation of the three uncommonness types related to the set of ideas of Table 2. The flags indicate the particular idea code. Bigger balls indicate ideas generated by both groups. Smaller balls indicate ideas generated only within a single group.

The suggested graphical representation allows to extract important considerations about the outcomes of an experiment. Indeed, for the illustrative example shown in Table 2, first of all it is important to observe that overall uncommonness and weighted uncommonness coincide only in three cases (the three balls on the diagonal). Therefore, it is evident that different assessment approaches often lead to different results, even if using the same attribute codification. Let us consider for example the mean values of the novelty scores (calculated on the two sets of thirty concepts shown in Table 2) obtained with both overall uncommonness and weighted uncommonness. As shown in Table 3 (neglecting statistical considerations for this illustrative example), it is quite evident that if following only the overall uncommonness approach, substantially no meaningful differences between the two groups can be observed in terms of mean scores.

Table 3. Mean novelty scores for the illustrative set of Table 2

	Mean (Group 1)	**Mean (Group 2)**	**Difference between groups**
Overall uncommonness	9,0	8,7	0,3
Weighted uncommonness	8,0	6,4	1,6

Moreover, the graph shown in Figure 1 provides an overall overview of the ideas that are generated by both groups (bigger balls). Consequently, the evaluator can extract interesting considerations immediately, and with the same codification procedure. For instance, Concept 321 in Group 2 (lower-left angle in Figure 1) gets a quite low score with both the metrics. This means that Group 2 used it quite frequently (low overall-uncommonness), and that PP3 and WP32 were used quite frequently too (low weighted uncommonness). The latter observation can be easily verified by observing ideas immediately above in the graph (i.e. 322, 323), whose particular combinations are more infrequent but are made with the same commonly used PP and WPs (for Group 2).

Since Idea 321 is represented with a bigger ball, it means that also Group 1 generated it but in this case it is located in the upper right angle of Figure 1. Therefore, while this idea and its PP and WP are used frequently in Group 2, it is one of the most infrequent combination in Group 1, where also PP 3 and WP 32 are rarely used. These observations can lead to important considerations, e.g. about the different background, expertise, fixation, etc. characterizing the two groups. For instance, the different distributions of balls in Figure 1, as well as the analysis of Table 2, reveal that Group 2 tends to fixate on PP 3, while Group 1 presents fixation on PP 1, but with a minor intensity. However, depending on the specific experimental objectives, other observations can potentially performed through the concurrent exploitation of the three-dimensional views of uncommonness.

5 DISCUSSION

Even if the set of pros and cons shortly reported in Table 1 can be useful for selecting the most suited assessment approach, it also revealed that SNM-based metrics and overall infrequence (Jansson and Smith, 1991; Linsey *et al.*, 2011) have a similar nature but different properties. Indeed, the overall infrequence allows to consider ideas in their entirety thus better highlighting unique occurrences. It means that unique concepts will get the highest possible score (a single concept in a bin), independently on the level at which the "uniqueness" appears. This may be useful or not, depending on the experiments' objectives and a sound discussion about that cannot be provided at the moment. Differently, it is easy to observe that the SNM-based metrics are more restrictive, and they can highlight unique concepts only if constituted by unique ideas (i.e. ideas appearing only one time) for each attribute. Indeed, since SNM-based metrics perform a weighted calculation of uncommonness, if even a single attribute is implemented by a non-unique idea, the overall concept cannot get the highest score. Nevertheless, if on the one hand this could be seen as a limitation, on the other hand it is a direct consequence of one of the advantages of SNM-based metrics. Indeed, they allow to extract novelty scores by taking into account the different weights assigned to attributes. In particular, GT versions have the double advantage to reduce subjectivity in the identification of attributes, and to consider the different impact of abstraction levels. Additionally, notwithstanding some non-negligible issues highlighted in Section 3, a further view is provided by the proposal of Lopez-Mesa and Vidal (2006), who present an interesting way for evaluating uncommonness across groups.

According to the different views offered by the reviewed metrics and related procedures, we proposed an integrated three-dimensional assessment approach. More precisely, the proposal aims at exploiting the advantages provided by SNM-family (in the example we referred to Peeters *et al.* (2010)) to reduce subjectivity in overall uncommonness assessments. Besides the weighted uncommonness normally calculated for SNM-based metrics, in our proposal also the overall uncommonness is calculated in terms of attributes' solutions constituting each idea, i.e. by considering their combinations. This is a less subjective way for creating the bins required in the original approach described in Section 2.1. Moreover, the same attribute-based codification is used to observe uncommonness of ideas across groups or teams. Then, if multiple groups are considered, it is possible to count how many groups share each particular concept and/or the ideas at different abstraction levels (thanks to the weighted uncommonness dimension). However, different requirements may characterize each specific design experiment but, as shown in the example reported in Section 4, the three-dimensional view and the proposed graph offer the possibility to perform a wide range of considerations in an easier way.

Concluding, this paper contributes to a better understanding of the multiple concepts of uncommonness, thus paving the way to future research activities about novelty assessment. The most important limitation of this work is the absence of a real application of the proposed approach, which is needed to obtain information about its applicability and usefulness. Nevertheless, this limitation directly suggests future research activities, aimed at the application of the proposal in real experimental contexts.

About the possible impacts of this work for industry, it is currently difficult to understand the context in which uncommonness could provide useful information and/or efficient tools. Indeed, notwithstanding the importance that firms attribute to innovation and then "also" to the novelty of their products, it is not trivial to find an industrial application field for a-posteriori novelty. This is another important research hint, which can be pursued by starting from the contents of this paper.

6 CONCLUSIONS

Starting from a literature review focused on the different procedures for a-posteriori novelty assessment, this paper provided a first comparison of those currently available. Highlighted differences revealed that the metrics offer different views of uncommonness, each of them potentially useful for experimental purposes. However, it has been highlighted here that relying only on a specific metric can lead to neglect important considerations about the outcomes of experimentations. Therefore, we proposed an integrated assessment approach where each dimension, explored by the reviewed metrics, can be evaluated concurrently. This paper is expected to help researchers in performing data analysis, and to provide useful hints for future research in the field of a-posteriori novelty assessment.

REFERENCES

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.

Corazza, G.E. and Agnoli, S. (Eds.). (2016), "Creativity in the Twenty First Century Multidisciplinary Contributions to the Science of Creative Thinking", Springer Science+Business Media Singapore, available at: https://doi.org/10.1007/978-981-287-618-8.

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.

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.

Koh, E.C.Y. and De Lessio, M.P. (2018), "Fixation and distraction in creative design : the repercussions of reviewing patent documents to avoid infringement", *Research in Engineering Design*, Springer, London, Vol. 0 No. 0, p. 0.

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.

Lopez-Mesa, B. and Vidal, R. (2006), "Novelty Metrics in Engineering Design Experiments", *9th International Design Conference, DESIGN 2006*, pp. 557–564.

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. (2008), "Studying engineering design creativity - Developing a Common Definition and Associated Measures", in Gero, J.S. (Ed.), *NSF International Workshop on Studying Design Creativity'08*.

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.

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*.

Tsenn, J., Atilola, O., McAdams, D.A. and Linsey, J.S. (2014), "The effects of time and incubation on design concept generation", *Design Studies*, Elsevier Ltd, Vol. 35 No. 5, pp. 500–526.

Vargas-Hernandez, N., Okudan, G.E. and Schmidt, L.C. (2012), "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. (2013), "Systematic Ideation Effectiveness Study of TRIZ", *Journal of Mechanical Design*, Vol. 135 No. 10, p. 101009.

Vargas-Hernandez, N., Shah, J.J. and Smith, S.M. (2010), "Understanding design ideation mechanisms through multilevel aligned empirical studies", *Design Studies*, Elsevier Ltd, Vol. 31 No. 4, pp. 382–410.