INTERNATIONAL DESIGN CONFERENCE - DESIGN 2012
Dubrovnik - Croatia, May 21 - 24, 2012.

ORIGINALITY AND NOVELTY: A DIFFERENT UNIVERSE

P. -A. Verhaegen, D. Vandevenne and J. R. Duflou

Keywords: design evaluation, effectiveness metrics, originality, novelty

1. Introduction

Although a plethora of idea/concept generation methods exist, their effectiveness is mostly claimed based on anecdotal evidence, or degree of educational and industrial applications [Tomiyama et al. 2009]. However, the steering of the development of ideation methods demands a structured analysis of the methods' effectiveness, as well as the mechanisms underlying their effectiveness [Hernandez et al. 2010]. Such results can also give a clear view on the correct circumstances under which to use certain ideation methods, e.g. a novel versus variant design task, or expert versus novice designers.
Insight in the effectiveness of an ideation method can be provided following two approaches [Shah et al. 2000], the process-based and the outcome-based approach. The former approach is based on the observation of specific cognitive processes in controlled experiments, but is considered time-consuming and subjective, while exhibiting a low external validity when extrapolating the results to more complex engineering problems [Shah et al. 2000]. For these reasons, the outcome-based approach, in which the evaluation of the effectiveness of a method on the generated designs themselves, is predominantly applied [Shah et al. 2000]. In outcome-based studies, different aspects of the generated designs, as well as the idea space formed by these designs, are quantified by a number of metrics, such as the quantity, the quality, the variety and the novelty [Shah et al. 2003]. Based on [Dean et al. 2006] and [Shah et al. 2000], Figure 1 overviews the different metrics and their subdimensions including different namings thereof.
The **quantity metric** is a commonly agreed on metric [MacCrimmon and Wagner 1994], measuring the number of unique generated concepts. In the context of this research, a concept is defined as a set of ideas (each from a different abstraction level), and their relationships, which are as a whole expected to solve one or more required functions as subject of an identified problem. A concept can be regarded as a coherent solution represented in sketches or wordings by the designer, e.g. a sketch of a new type of razor. The **variety metric** proposed by [Shah et al. 2000] is a formal measure of the explored design space operating on a tree-encoded structure of the idea space, and is to be interpreted as a metric for the shape of the idea space tree. A low variety indicates a slim tree with closely related concepts, while a high variety indicates a wide tree with concepts differentiated at high abstraction levels, i.e. concepts using different physical or working principles. The **quality metric** as proposed by [Dean et al. 2006] encompasses dimensions as workability, relevance, specificity and novelty, although the latter is mostly treated as a metric in its own right [Connolly et al. 1990], [Jansson and Smith 1991], [Shah et al. 2000], [Linsey 2007], [Srinivasan and Chakrabarti 2010]. The **novelty** can be further subdivided in originality and paradigm relatedness [Dean et al. 2006]. The former expresses the degree to which a concept is not only rare, but is also ingenious, imaginative, or surprising, while the latter expresses the degree to which a concept is radical or transformational. Since the rarity of an concept can more easily be quantified, other research explicitly reports the rarity [Connolly et al.

1990], uses it as a substitute for originality [Jansson and Smith 1991], or as substitute for the novelty metric as a whole [Shah et al. 2000], [Linsey 2007]. Recently, a number of studies have used outcome based metrics to prove and compare the effectiveness of ideation methods [Linsey 2007], [Peeters et al. 2010], [Chan et al. 2011], to delineate proper usage conditions [Linsey 2007], [Peeters et al. 2010] and to investigate the effectiveness of underlying mechanisms [Hernandez et al. 2010]. However, rarity calculations are not consistent between studies. The contribution of this paper is to highlight this inconsistency, and to link it to different approximations of the *universe of ideas for comparison*.
This remainder of the paper is structured as follows. Section 2 overviews originality and novelty metrics. Section 3 illustrates the circumstances under which the applied approximations in the rarity calculations are likely to lead to errorneous results. The final section formulates the conclusions.

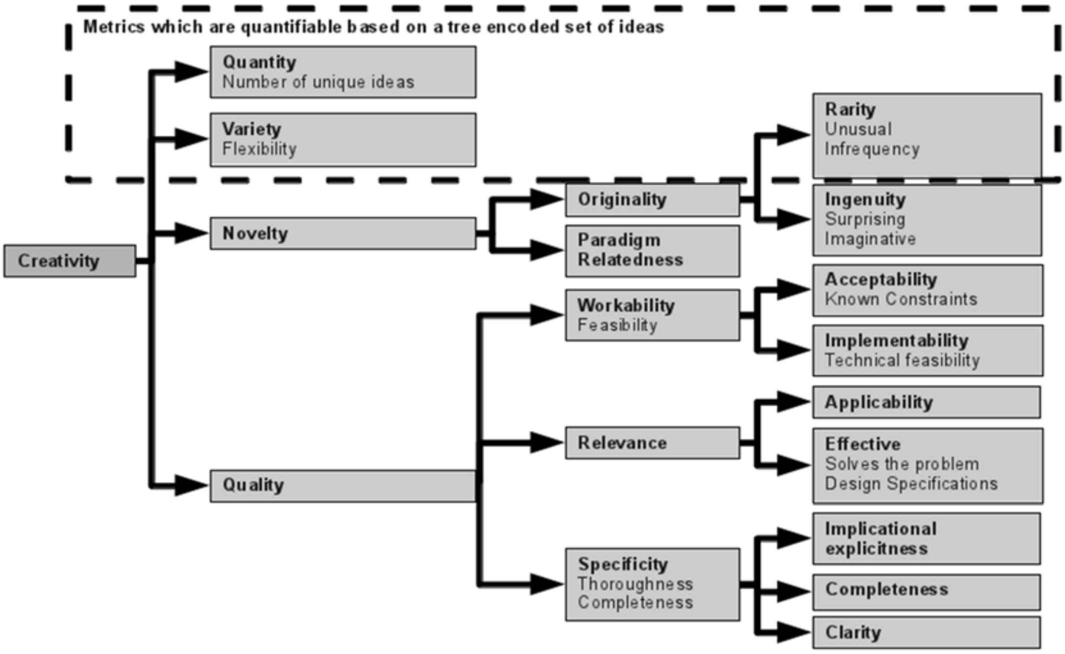

Figure 1. Overview of the ideation effectiveness metrics

2. Originality and novelty

Both dimensions of novelty, originality and paradigm relatedness, can only be interpreted in respect to an unknown universe of known concepts to compare against, i.e. a concept can only be considered rare, ingenious, imaginative, surprising, radical or transformational in respect to a set of known solutions. This universe is called the *universe of ideas for comparison (UnIC)* [Shah et al. 2000].
For an ideation exercise, the UnIC can be approximated following two approaches. In a first approach, the UnIC is approximated by the concepts generated by the participants themselves. In this case, the UnIC can be used to calculate the rarity of concepts (and attributes thereof) in the set of generated concepts. Another technique to calculate the novelty is to implicitly form the UnIC by relying on expert opinions. The experts estimate the originality or paradigm relatedness of concepts based on existing solutions or attributes of existing solutions known to them. The interrater reliability of the expert opinions is to be determined, indicating, but not guaranteeing, a good choice of domain experts and hence a reliable novelty/originality estimation.
The sections below cover a number of rarity related metrics, of which Table 1 gives an overview based on the main differences. The first distinction is based on the level used to assess the novelty, originality or rarity of a concept, e.g. Shah et al. [Shah et al. 2003] propose a novelty metric calculated based on the novelty of the attributes of a concept. The second distinction made in Table 1 is the applied UnIC approximation discussed hereabove, i.e. the approximation of the UnIC based on expert knowledge versus concepts generated by the participants.

Table 1. Overview the main differences between rarity and related metrics

		Applied UnIC approximation	
		Expert knowledge	Generated concepts (rarity)
Level used to assess novelty, originality or rarity of a concept	Concept	*Originality* (Hood) *Creativity* (Amabile, Caroff)	*Originality* (Jansson & Smith, Cropley, Guilford, Torrance)
	Attribute of a concept	*Novelty* (Shah first approach)	*Novelty* (Shah second approach)

2.1 Originality

Early work by Jansson and Smith [Jansson and Smith 1991] investigates the effect of providing designers examples prior to a design exercise. Half of the designers were subjected to a priming example, while the remainder constituted the control group. Metrics were calculated to prove that design fixation did occur as a result of the exposure to a design example. The authors proposed an *originality* metric computed as the average 'o' score for an individual's concepts divided by the number of concepts for that subject. The 'o' score for each item was calculated as *o=1-((number of similar designs generated by other subjects)/(total number of designs for all subjects))*. Based on the taxonomy of Figure 1, the originality metric proposed by Jansson & Smith is in fact a *rarity* metric, calculated by comparing each concept to the concepts generated within the group subject to the same experimental conditions.

Since Jansson & Smith use two conditions (fixation and control), the concepts are compared to two approximations of the UnIC. However, taken into account the conclusions drawn, the calculated originality is interpreted in respect to the other participants within the same experimental condition as to provide evidence of design fixation. Hence, the use of two different approximations of the UnIC can be considered legitimate.

The originality metric proposed by Jansson & Smith [Jansson and Smith 1991] is conceptually similar to the originality metrics proposed by [Torrance 1966], [Guilford 1967], [Cropley 1967] in the area of creative and divergent thinking, e.g. Guilford rates the originality of an concept based on the rarity; a concept is awarded one or two points if the occurance frequency of similar concepts among all responses is less then 5, respectively, one percent. The main similarity between these originality metrics is that they are calculated based on a UnIC comprised of the concepts generated by the participants themselves. The definition of the UnIC can be adapted in the originality metrics described in Section 2.1, e.g. [Cohen 1972] applied Cropley's algorithm and established scoring categories for each originality item based on a UnIC comprising concepts from 100 random students throughout Australia.

Amabile [Hennesey and Amabile 1999] proposed the Consensual Assessment Technique using raters at a concept level to evaluate the more encompassing creativity metric, of which Caroff [Caroff 2008] identified the originality and appropriateness as the two main components. Caroff also showed that the creativity judgments of raters depends on the creative ability of the raters themselves, confirming the conclusions drawn by [Hood 1973] for raters assessing the originality of concepts. These results also show that expert raters do not tend to differentiate between the concepts at a low and medium originality level, in turn questioning whether expert raters should be used at all.

2.2 Novelty

Similar to the above described originality metrics, Shah et al. [Shah et al. 2000] propose the novelty metric to measure ideation effectiveness. According to Shah et al. the novelty is intended to measure the unusualness of a concept, which translates to the rarity metric in Figure 1. A concept's novelty is based on occurence frequencies, or expert knowledge, of the *attributes* of the generated concept, and not on the rarity of the *total* concept as in the case of the originality metrics as illustrated in Table 1. Shah et al. [Shah et al. 2000] explicitly define two approaches to estimate the UnIC. In the first approach, the UnIC is approximated by all preconceived concepts assembled by experts before analyzing the results of the experiments. In practice, experts define one or several attributes on which to assess the unusualness of the generated concepts based on the specifics of the design problem, e.g.

physical principle used to remove hair from the face for a new razor design problem, or motion exhibited by the razor blades. The second approach proposed by [Shah et al. 2000] is to approximate the UnIC by the set of all concepts generated by the experiment participants themselves. Shah et al. [Shah et al. 2003] further demonstrate that the second approach allows to automatically calculate the rarity of a design based on the occurrence frequency of attributes of the design in the complete set of designs generated by the participants to an experiment. Although calculated based on occurrence frequencies of attributes of concepts, this approach is similar to the originality proposed by Jansson & Smith [Jansson and Smith 1991], [Guilford 1967], [Cropley 1967] and [Torrance 1976] since all methods define the UnIC to be comprised of the concepts generated by the participants themselves.

3. Originality calculations in priming ideation methods

The above sections illustrated several methods to estimate the originality and rarity of concepts based on different approximations of the UnIC and different levels of detail in the frequency calculation, i.e. occurrence frequencies at the concept or attribute level. The appropriateness of the approximation of the UnIC is to be evaluated for each research study depending on the circumstances and the conclusions to be drawn from the results.

For the UnIC approximation by **expert knowledge**, the originality estimations can be considered dependant on the choice of experts [Hood 1973], [Caroff 2008], although considerable level of interrater agreements have been reported in practice for the creativity [Hennessey and Amabile 1999], of which a main dimension is the originality [Caroff 2008]. However, as detailed by the paragraphs below, the rarity calculations based on the UnIC approximation by **the generated concepts** are also often miscalculated for priming ideation methods.

In priming ideation methods, the participants are subjected to stimuli, e.g. existing solutions, intended to enhance certain aspects of the designs, such as the novelty or technical feasibility, it can be expected that the participants are fixated on certain solutions or attributes of solutions. As detailed in Section 2.1, Jansson & Smith's prove that the rarity (reported as originality) of an idea space formed by primed participants can be lower than the rarity of the idea space formed by the concepts generated by non-primed participants (control group), in turn proving the effect of design fixation by design examples. It follows that the concept space formed by the fixation group should not be used as a basis to approximate the UnIC, and hence that the originality is not to be interpreted as an originality in respect to all known solutions.

Similarly to the above described research by [Jansson and Smith 1991], a number of studies have approximated the UnIC by concepts generated by a primed group of participants. Following the reasoning explained above, the generated designs are not drawn from the UnIC at a frequency reflecting the actual (but unknown) occurrence frequency of these designs. Instead, the frequency at which solutions with certain attributes or functions are generated is now dependent on the stimuli given to the participants. It is therefore unlikely that the set of designs, generated by primed participants, is a good representation of the UnIC, and unlikely that the calculated novelty score of an attribute truthfully reflects the unknown novelty score of the attribute in the UnIC. However, in contrast to Jansson & Smith, which have only drawn conclusions from the rarity within a group or condition as to prove the effect of fixation, these studies often draw conclusions on the rarity in respect to all known solutions.

The following example illustrates the effect on the rarity calculation of including primed participants in the approximation of the UnIC. Table 2 gives possible outcomes for a hypothetical ideation session. The first three columns give the number of generated concepts for both possible values of the *motion* attribute, e.g. the control group generated three concepts which exhibit a rotating motion, and two concepts having sliding as motion type. The rarity calculations are performed according to the methodology proposed by Shah [Shah 2003], e.g. the rarity of the rotate motion type for the control group equals the number of concepts not having this motion type divided by the total number of generated concepts, or (5-3)/5 = 0.4. From Table 2, it can be seen that the scores for an attribute rarity score are dependant on the number of concepts exhibiting different attributes. Hence, a concept exhibiting rotating motion will have a rarity score of 0.4, 0.2 or 0.6 depending on whether the UnIC is approximated by the concepts generated from, respectively, the control group, the group fixated with example 1 or the group fixated with example 2 (another option would be to calculate the rarity based on the concepts generated under all conditions). In the extreme case, a method targeted at ideating designers to develop concepts with only a rotating motion will lead to concepts with a rarity score of zero.

Table 2. Number of concepts generated and resulting rarity score for control and primed conditions

		Number of generated concepts			Rarity calculation		
		Control no fixation	Fixated groups		Control no fixation	Fixated groups	
			example 1	example 2		example 1	example 2
Motion type	Rotate	3	4	2	(5-3)/5=0.4	0.2	0.6
	Slide	2	1	3	(5-2)/5=0.6	0.8	0.4
Sum		5	5	5			

The above example illustrates a possible flaw in the calculation of the rarity score for primed participants based on Shah et al.'s metric utilizing the attribute occurrence frequencies. However, since the example only uses a single attribute (motion type), the example degenerates to the case in which concepts are grouped as similar by experts on a concept level. Furthermore, although the conclusions are drawn from a single attribute example, they remain valid for the case in which concepts are grouped based upon multiple attributes. To resolve this, it is proposed to explicitly limit the concepts on which to base the calculation of the attribute and concept rarity scores to the concepts generated by participants which are believed to generate concepts with an occurrence frequency resembling to the UnIC. Concepts generated by primed participants are to be excluded from the calculation of the attribute or concept category rarity scores, i.e. only concepts from non-primed participants are appropriate in order to calculate the total number of concepts and the number of similar concepts, denoted by respectively T_{jk} and C_{jk} in [Shah et al. 2003].

Although the above mentioned solution has occasionally been applied by researchers in the field of design creativity [Wilson et al. 2010], [Peeters et al. 2010], [Chan et al. 2011], none of these studies explicitly detail the reasons for restricting the approximation of the universe of ideas for comparison (UnIC) to the designs generated by non-primed participants. Also, other studies [Linsey et al. 2011], [Neeraj 2010], [Wodehouse and Ion 2011] do not restrict the UnIC to the designs generated by non-fixated or non-primed participants, or do not mention this in the reported results. Based on the reasoning detailed above, it can be expected that these studies under- or overestimate the rarity (often reported as originality or novelty) of the applied priming ideation methods.

4. Summary

Over the recent years, a number of studies have applied effectiveness metrics, such as the variety or novelty, on the outcome of ideation exercises to compare ideation methods. The rarity metric, sometimes reported as a substitute for the originality or novelty, is often calculated based on concepts generated by the participants to the ideation exercise themselves. Depending on the methodology the generated concepts can be clustered as similar on concept level, or grouped along different attributes of the generated concepts, e.g. type of motionary action. It was shown that the calculated rarity of concepts (and attributes) is dependant on the concepts used to approximate the *universe of ideas for comparison*. Based on a hypothetical example, it was also shown that the inclusion of concepts from primed participants can lead to a errorneous rarity calculation. The proposed solution follows directly from this observation, and entails the exclusion of the concepts generated by primed participants.

Although the above mentioned solution has been occasionally applied in studies, none have explicitly detailed the reasons for restricting the approximation of the universe of ideas for comparison to the designs generated by non-primed participants. Furthermore, a number of studies do not restrict the universe of ideas for comparison to the designs generated by non-fixated or non-primed participants, or do not mention this in the reported results. It was shown that the latter studies are expected to under- or overestimate the actual rarity score, in turn misrepresenting the originality or novelty of the applied ideation method. It can be expected that a misrepresentation of an ideation method's effectiveness will in turn lead to errorneous conclusions concerning the applicability of the ideation method, and mislead the future development efforts of ideation method research. It is therefore important for future ideation method effectiveness studies to clearly state which concepts make up the applied approximation of the universe of ideas for comparison.

References

Caroff, X., Besançon, M., "Variability of creativity judgments. Learning and Individual Differences", Vol. 18, No. 4, 2008, 367-371.

Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., Kotovsky, K., "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples", Journal of Mechanical Design, Vol. 133, No. 8, pp. 081004.

Cohen, D., "The development of a creativity test", Research in Science Education, Vol. 2, No. 1, 1972, 82-95.

Connolly, T., Jessup, L. M., Valacich, J. S., "Effects of anonymity and evaluative tone on idea generation in computer-mediated groups", Management Science, Vol. 36, No. 6, 1990, pp. 689-703.

Cropley, A. J., "Creativity", Longman, London, 1967.

Dean, D. L., Hender, J. M., Rodgers, T. L., Santanen, E. L., "Identifying quality, novel, and creative Ideas: Constructs and scales for idea evaluation", Journal of the Association for Information Systems, Vol. 7, No. 10, 2006, pp. 646-698.

Guilford, J. P., " The Nature of Human Intelligence", McGraw-Hill Book Company, New York, 1967.

Hennessey, B. A., Amabile, T. M., "Consensual assessment", Encyclopedia of creativity, Runco M. A., Pritzker S. R. (Eds.), Academic Press, San Diego, California, 1999.

Hood, R. W., "Rater Originality and the Interpersonal Assessment of Levels of Originality", Sociometry, Vol. 36, No. 1, 1973, pp. 80-88.

Jansson, D. G., Smith, S. M., "Design fixation", Design Studies, Vol. 12, No. 1, 1991, pp. 3-11.

Linsey, J. S., "Design-by-Analogy and Representation in Innovative Engineering Concept Generation", PhD Thesis, The University of Texas, Austin, Texas, 2007.

Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., Markman, A. B.,, "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, 2011, pp. 031008.

MacCrimmon, K. R., Wagner, C., "Stimulating Ideas Through Creativity Software. Management Science", Vol. 40, No. 11, 1994, pp. 1514-1532.

Neeraj, K., " Comparative Study of Function Modeling Methods using Protocol Analysis". Thesis, Texas Tech University, Texas, 2010.

Peeters, J., Verhaegen, P.-A., Vandevenne, D., Duflou, J. R., "Refined Metrics for Measuring Novelty in Ideation", Proceedings of the IDMME - Virtual Concept 2010 - Research in Interactive Design, Bordeaux, Vol. 3, 2010, pp. 127.

Shah, J. J., Kulkarni, S. V., Vargas-Hernandez, N., "Evaluation of Idea Generation Methods for Conceptual Design: Effectiveness Metrics and Design of Experiments", Journal of Mechanical Design, Vol. 122, No. 4, 2000, pp. 377-384.

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

Tomiyama, T., Gu, P., Jin, Y., Lutters, D., Kind, C., and Kimura, F, "Design methodologies: Industrial and educational applications," CIRP Annals-Manufacturing Technology, Vol. 58, No. 2, 2009, pp. 543-565.

Torrance, E. P., "The Torrance Tests of Creative Thinking – Norms-Technical Manual", Personnel Press, Princeton, New york, 1966.

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

Wilson, J. O., Rosen, D., Nelson, B. A., Yen, J., "The effects of biological examples in idea generation", Design Studies, Vol. 31, No. 2, 2010, 169-186.

Wodehouse, A., Ion, W., "Augmenting the 6-3-5 method with design information", Research in Engineering Design, Vol. 133; No. 3; 2011, pp. 031008.

Ir. Paul-Armand Verhaegen
Katholieke Universiteit Leuven
Celestijnenlaan 300A Bus 2422
3001 Heverlee, Belgium