

Enhancing novelty with knowledge-based support for Biologically-Inspired Design



D. Vandevenne, T. Pieters and J. R. Duflou, Centre for Industrial Management, Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300A, 3001 Leuven, Belgium

With a two-decade consistent research interest for Systematic Biologically-Inspired Design, a number of methods and tools to support bio-ideation have been proposed. However, objective quantification of the effects these aids have on the design outcomes is rare. This contribution presents an impact analysis of the most popular knowledge-based tool, AskNature, in the form of an outcome-based study. The results consistently support a common claim used in favour of bio-inspired design, i.e. the expectation of identifying more out-of-the-box solutions. Furthermore, to further facilitate biological solution analysis and cross-domain knowledge transfer, an adaptation to AskNature's stimuli format — i.e. adding a graphical illustration of the biological solution principle — is validated to further boost novelty.

© 2016 Elsevier Ltd. All rights reserved.

Keywords: conceptual design, creative design, design method(s), evaluation, Biologically-Inspired Design

Creative problem solving is a key task for companies pursuing inventions that may grow into successful innovations. One strategy for solving new problems is learning from previously solved analogous problems. In Design-by-Analogy (DbA) a solution principle behind an already solved problem is transferred to solve a new problem. For example, when looking for new ways to unfold a tent, products with similar functionality, like for instance umbrellas, can be sources of inspiration and knowledge transfer. Biologically-Inspired Design (BID) is a specific type of between domain DbA where inspiration is taken from the natural world (source domain) to solve technical problems or challenges (target domain). Three frequently used arguments for looking at nature for inspiration are (1) the proven performance of biological systems, (2) the potential for sustainable products and (3) the potential for finding out-of-the-box solutions. The first, proven performance of biological systems, logically follows from the overwhelming evidence supporting evolution, i.e. the change in inherited characteristics of biological populations over successive generations, a continuous repetition of a non-random selection mechanism (survival of the fittest) applied to traits subject to random variation. The products of these continuous improvement iterations — solution principles of biological systems —

Corresponding author:
Dennis Vandevenne
dennis.vandevenne@kuleuven.be



www.elsevier.com/locate/destud
0142-694X *Design Studies* 46 (2016) 152–173
<http://dx.doi.org/10.1016/j.destud.2016.05.003>
© 2016 Elsevier Ltd. All rights reserved.

offer attractive starting points compared to an empty drawing board. The second and third arguments, however, are currently not backed by adequate empirical evidence. Although it seems logical that a natural world, where problems are solved with renewable material and energy resources while avoiding the creation of ever-growing waste piles, has the potential for inspiring (more) sustainable technical solutions, systematic evidence quantifying this popular advantage currently is missing. The third argument, enhanced potential for finding out-of-the-box solutions, is motivated by the fact that biology is a largely untapped knowledge domain and that bio-inspired design uses far analogies with little common ground in superficial properties. Although previous studies measured the impact of following a course in bio-inspired design (Nelson, Wilson, & Yen, 2009), of presenting a single biological solution versus a technical solution (Wilson, Rosen, Nelson, & Yen, 2010) and of directing students to think of nature for problem solving (Glier, Tsenn, and McAdams (2012), currently no knowledge-based bio-ideation tool has been quantitatively evaluated. Therefore this contribution measures the effect the use of a knowledge-based BID tool has on the generated design concepts. For this purpose the impact of the most commonly used tool, AskNature, is quantified with an hereto designed outcome-based experiment. Furthermore, a potential improvement of AskNature's standard stimuli representation is investigated.

The next section provides the reader with the necessary background in Systematic Biologically-Inspired Design (SBID), outcome-based validation metrics and of the state-of-the-art in BID outcome-based experiments. Section 2 details the experiment's hypotheses, Section 3 the experiment setup and Section 4 the results. Finally the discussion of the results and general conclusion are respectively presented in Sections 5 and 6.

1 Background

First, existing bio-ideation tools are listed and the choice of AskNature is motivated in Section 1.1. Next, the four typical metrics used to quantify an ideation tool's performance are detailed in Section 1.2. These are quantity, variety, novelty and quality (Shah, Smith, & Vargas-Hernandez, 2003). Finally, Section 1.3, discusses findings from previous BID experiments and extracts opportunities for improvement of the proposed experimental design.

1.1 BID supporting tools

The last two decades a research community grew that aims at developing methods and tools to support the Systematic BID process. These approaches include keyword-based search methods (Kaiser, Farzaneh, & Lindemann, 2014; Lenau, Dentel, Ingvarsdóttir, & Gudlaugsson, 2010; Shu, 2010), approaches supporting on the classification of biological strategies (Deldin & Schuknecht, 2014; Vincent, Bogatyreva, Bogatyrev, Bowyer, & Pahl, 2006),

contributions that require complex model instantiation for each corpus entry (Chakrabarti, Sarkar, Leelavathamma, & Nataraju, 2005; Nagel & Stone, 2012; Vattam, Wiltgen, Helms, Goel, & Yen, 2010) and one recent approach based on technical patent and biological paper mining (Vandevenne, Verhaegen, & Duflou, 2015).

From these contributions, AskNature — developed by The Biomimicry Institute — is, by far, the most popular, reporting almost 1.8 million hits and 478 486 unique visitors in 2012 (Deldin & Schuknecht, 2014). Furthermore, a scalable extension is proposed to this approach by Vandevenne, Verhaegen, Dewulf, and Duflou (2014) who developed an algorithm aiming at automated classification of biological strategies in AskNature’s Biomimicry Taxonomy. Scalability, in this context, refers to the ability of the approach to integrate large biological knowledge bases. For these reasons, AskNature is chosen as the first knowledge-based bio-ideation tool to perform outcome-based experiments on and this paper reports the results hereof.

Next follows a short description of the use of AskNature positioned in the four typical phases of the simplified SBID process (Vandevenne, Verhaegen, Dewulf, & Duflou, 2011). To use AskNature, the designer formulates his or her problem (Step 1) into AskNature’s Biomimicry Taxonomy, a functional, three-level hierarchical classification mechanism developed to organize biological strategies (Deldin & Schuknecht, 2014). After problem formulation (Step 1), one or a couple of interesting functional categories are selected from the Biomimicry Taxonomy and a list of biological strategies is returned as search results (Step 2). Next, this list is analysed and filtered (Step 3) to retain a small number of interesting biological strategies for knowledge transfer (Step 4) to compose one or more cross-domain analogies expressed in technical concepts that address the challenge at hand. More details about AskNature, its use and its scalability are provided by Deldin and Schuknecht (2014) and Vandevenne et al. (2014).

1.2 Outcome-based validation

This section briefly discusses the state-of-the-art metrics used for outcome-based validation in this contribution. For a more elaborate discussion of these metrics and further argumentation of the selected metric variants see Chapters 3 to 5 of Verhaegen (2013).

1.2.1 Quantity

‘Quantity’ measures the number of unique concepts generated by a participant during a design challenge. To facilitate concept counting, the participants are instructed to separate concepts in different boxes on the answering forms and to provide a sketch as well as a short textual description. Incomplete concepts and concepts resulting from misunderstanding the challenge are omitted,

compound concepts are disaggregated and duplicate concepts are considered as one. These actions typically occur with low frequency. For example, the first brainstorm presented in Section 3.4 resulted in 313 concepts of which only one was omitted as there was no sketch and the one sentence provided was too vague to integrate the concept in a genealogy tree.

1.2.2 Variety

‘Variety’ measures to which extent and individual’s set of generated ideas or concepts differ from each other. A prerequisite for variety calculation is the positioning of the concepts in a genealogy tree (Shah et al., 2003), i.e. a hierarchical tree-like structure with four abstraction levels that organizes the concepts according to their similarities and differences. At the highest level of abstraction concepts are distinguished from each other by the underlying physical principle. At the second level, concepts can share the same physical principle, but are distinguished from each other based on their working principle. On the third and fourth level, the distinction is based on embodiment and details, respectively. In the context of such a genealogy tree, a concept is defined as a set of ideas — one for each abstraction level — and their relationships, and represents a coherent solution expressed in a sketch and the wordings of the designer (Verhaegen, 2013, p. 5). The lower abstraction levels (embodiment and detail) are sometimes omitted by the participants, e.g. when they are instructed to document their ideas with a rough sketch and just a few words for clarification.

An overview of the different approaches to variety calculation is offered by Verhaegen, Vandevenne, Peeters, and Duflou (2013); who propose two main refinements to the state-of-the-art. First, a level-based variety score is detailed to allow the evaluation of the idea set on each abstraction level in the genealogy tree. Hereby information is reported that potentially is lost when using an aggregated variety metric. Furthermore, Verhaegen et al. (2013) propose a refinement to the variety metric that takes into account the uniformity of the distribution of ideas. By integrating the inverse of the Herfindahl index (Herfindahl, 1950), the new proposed variety calculation demonstrates a more gradual behaviour and avoids overestimation of the actual variety of the idea set. These refinements taken into account, the level-based variety metric is calculated as follows:

$$V_k = 10 \sum_{j=1}^m f_j \left(\frac{1}{n \sum_{i=1}^{b_k} p_i^2} \right) \quad (1)$$

where V_k is the variety score for abstraction level k , b_k is the number of nodes in the genealogy tree at level k , p_i refers to the probability of node i , m is the

number of different functions, f_j reflects the weight corresponding to function j and n is the number of concepts in the set.

1.2.3 Novelty

'Novelty' is approximated by measuring rarity or the unusualness of concepts or ideas. The novelty of each concept (Shah et al., 2003) can be calculated according to

$$N = \sum_{j=1}^m f_j \sum_{k=1}^n S_{jk} p_k \quad (2)$$

where N expresses the overall novelty for a concept with m functions and n abstraction levels in the genealogy tree, the weights of which are, respectively, expressed by f_j and p_k . S_{jk} represents the novelty of the function j at abstraction level k , calculated as follows (Shah et al., 2003)

$$S_{jk} = 10 \frac{T_{jk} - C_{jk}}{T_{jk}} \quad (3)$$

where T_{jk} is the total number of concepts generated for function j and level k , while C_{jk} expresses the count of the current solution for function j and level k . The multiplication by 10 scales the novelty score to a [0–10] range, with 10 representing the highest possible novelty score.

A statement about novelty is only meaningful in reference to a universe of known solutions. This universe is named the Universe of Ideas for Comparison (UnIC) (Shah, Kulkarni, & Vargas-Hernandez, 2000) and is reflected by T_{jk} and C_{jk} in Equation (3). As this UnIC is unknown, it needs to be approximated. As demonstrated by Verhaegen (2013) (Chapter 4), for priming ideation methods — i.e. methods that present stimuli likely to result in fixation on certain solutions or attributes of solutions — the set of concepts from the non-primed control group should be taken as approximation of the UnIC. The concepts of the non-control groups are excluded from the UnIC as it is likely that their generated designs are not drawn at a frequency reflecting the actual occurrence of these designs in the UnIC (Verhaegen, 2013, Chapter 4, p 89). To bring novelty scores in the range of zero to ten, concept novelty scores are divided by the sum of the level-based weights; and to use this range more efficiently level-based novelty scores are rescaled according to Verhaegen (2013).

1.2.4 Quality

Different contributions have associated multiple dimensions to the quality metric, e.g. workability (acceptability and implementability), relevance (applicability and effectiveness) and specificity (thoroughness or completeness) (Dean, Hender, Rodgers, & Santanen, 2006; Verhaegen, 2013). This study

follows recent studies (Glier et al., 2012; Kim, McAdams, & Linsey, 2014; Verhaegen, 2013) that take a well-defined three-point rating scale (Linsey, 2007) to assess the technical feasibility or implementability aspect of quality. For each concept a technical feasibility score is determined by answering the two questions depicted in Figure 1.

1.3 BID outcome-based validation studies

In this section, an overview of previous outcome-based bio-ideation experiments is presented. These studies measure the impact of the proposed method or tool with the in Section 1.2 described metrics. The results of the studies are discussed and opportunities for improvement of the experimental approach are extracted.

Nelson et al. (2009) measured the effect of participation in a BID course. A group of 11 students of Mechanical engineering – volunteers in a design class with special focus on BID – was compared to a control group of 26 students of a capstone mechanical engineering design class. One handpicked design challenge was given to the participants as focus for the experiment. The group of BID students demonstrated a variety score that was 37% higher than the control group and a 80% novelty increase. The authors report that none of the BID students actually used a BID solution principle. One possible explanation given is self-selection, i.e. the attraction of the BID-course of more motivated, interested and creative students. Another potential explanation is given as the explicit focus on innovation and Design-by-Analogy during the BID-course (Nelson et al., 2009). Without random test groups or pre- and post-BID course performance quantification, the reported performance improvements cannot be addressed to BID with certainty.

Wilson et al. (2010) quantified the effect of exposure to a biological example and compared the results against the use of a human engineered example and a control condition without an example. Hence, this experiment presents one handpicked design problem and measures the impact of one biological and one technological solution stimulus compared to a no stimulus condition. The

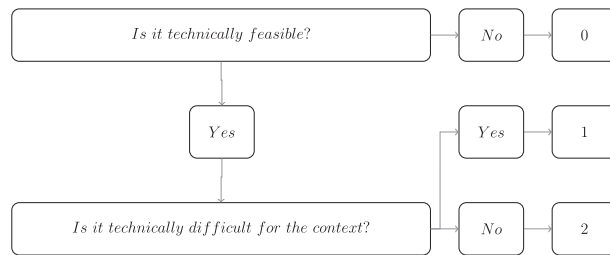


Figure 1 Applied quality scale (measuring technical feasibility) (Linsey, 2007)

authors conclude that exposure to the biological example increases novelty without reducing variety.

[Glier et al. \(2012\)](#) tested the impact of using the directed method — i.e. directing participants to consider how nature would solve the problem, without giving specific biological stimuli — on the outcomes of an ideation effort. Two different technical challenges are chosen for their controlled experiment. The authors report no significant effect on any metric (quantity, variety, novelty, quality) compared to a control condition where no such direction is given. This result is not surprising as engineers typically lack active knowledge of biological solution principles which severely limits their capacity for identifying cross-domain analogies. This is the reason why the tools, discussed in [Section 1.1](#), have been and are being developed. The reported results further advocate continued research on (Scalable) Systematic Biologically-Inspired Design and the authors explicitly state impact quantification of existing tools as relevant future work.

[Kim et al. \(2014\)](#) investigated the impact of the presentation format, i.e. one or multiple stimuli on a page, and valuableness of biological stimuli, i.e. their ability to trigger ideas for the problem at hand. A total group of 75 Mechanical Engineering students was randomly divided into six groups, two control groups and four groups that received different biological passages, represented by single lines from a corpus of biological papers. As was the case for most of the above studies, one handpicked design challenge forms the subject of the experiment. Novelty was not affected by any test condition, variety is significantly affected by valuableness of passages and presentation format. The authors conclude that high quality passages are preferred and they should be presented with multiple per page ([Kim et al., 2014](#)).

[Durand et al. \(2015\)](#) recently compared five different BID methods (Directed, BioTRIZ, functional modelling, case study and AskNature) on the dimensions of quantity, quality, student self-efficacy and student feedback. Compared to the above contributions, variety and novelty are not measured. The authors find encouraging results for both quality and quantity as well as for students design confidence, outcome expectancy and motivation. Interestingly the students preferred BioTRIZ while this method did not perform better in the analyses.

By analysing the above reference BID outcome-based studies, a number of opportunities are exposed for improvement of the experimental setup presented in [Section 3](#). First, three out of five of the above studies take one design case as basis for the tests. It is however a reasonable assumption that the measured impact of a BID course, biological or technological stimuli hypotheses and bio- or non-bio-ideation tools fluctuates for different design challenges and that a number of such tests needs to be conducted to draw more general

conclusions. Second, all of the above priming studies adopt handpicked design challenges. Ideally, the selection of test cases should be random, or at least not influenced by the current study, chosen stimuli, method or tool. Third, after random group assignment, group balance with respect to the adopted metric(s) needs verification by analysis of performance on a first, independent design challenge. Fourth, none of the previous experiments measures the impact on novelty of a knowledge-based ideation tool currently used by the community for bio-ideation. Section 3 details the experimental setup that takes into account these identified opportunities, but first the experiment's main hypotheses are formulated in Section 2.

2 Hypotheses

When analogies are based on solutions from a different source domain than the target domain they are often associated with more original, inventive analogies (Dahl & Moreau, 2002; Schild, Herstatt, & Lüthje, 2004; Tseng, Moss, Cagan, & Kotovsky, 2008; Ward, 1998). Indeed, by drawing inspiration from a largely untapped biological knowledge domain, using far analogies, a reasonable expectation is an enhanced probability of generating ideas that can lead to leapfrog innovations (Gebeshuber, Gruber, & Drack, 2009; Schild et al., 2004) and even new avenues of research (Fish & Beneski, 2014). Additionally, the growing number of innovative bio-inspired designs (Bonser, 2006) in diverse application domains (Bar-Cohen, 2011) provide extra expectations for finding out-of-the-box solutions. Hence, the main hypotheses focus on the novelty of the design outcomes.

The first and principal hypothesis states that exposing participants to the stimuli from a relevant functional category from AskNature increases novelty of the design outcomes. As AskNature typically only show photographs of the involved organisms, a second test condition replaces this photograph by a graphical illustration of the biological solution principle in each of the standard AskNature stimuli with as goal to positively impact the participants' ability to analyse, understand and transfer the strategy from nature. Hence it is expected that this adapted stimuli representation further increases novelty as more of the biological stimuli — which are expected to increase novelty by the first hypothesis — can be cognitively processed. The second hypothesis states that replacing the photograph of the organism with a graphical illustration of the biological solution principle further increases novelty compared to the standard AskNature representation.

Results for the other metrics will also be calculated and presented, as it is important to be aware of other potential advantages and disadvantages of exposing participants to biological strategies from AskNature's functional categories. Besides the above listed main hypotheses, other reasonable expectations are a drop in quantity (as processing biological stimuli takes time and

cognitive effort) and a drop in technical feasibility (as distant analogies might inspire harder to realise concepts).

3 Experimental setup

The outline of the experiment is illustrated by Table 1. The following subsections discuss the participants (Section 3.1), the three design cases (Section 3.2), test conditions and stimuli (Section 3.3), and experiment structure and time line (Section 3.4).

3.1 Participants

This experiment was conducted with 41 Master of Mechanical Engineering students at the University of Leuven (KU Leuven) and 18 students Master of Science in Industrial Design Engineering Technology at Ghent University. The students Mechanical Engineering from Leuven gained design experience in the course *Problem Solving and Designing*. This course consists out of a technical challenge given to teams of six people during one or two semesters. They also took a course named *Design Methodologies*, where teams of four students are assigned a comprehensive design task for a semester project. In this course they also learnt about the necessary theoretical background on design methodologies. Besides a general background in physics, chemistry, mathematics, etc, the Industrial Design Engineering Technology students focus on product design aspects like creativity, CAD, production techniques, prototyping, etc. The study is hands-on focussed with several practical design assignments, guiding students from problem identification to building prototypes. Students from the Master Mechanical engineering are taught more theoretic foundation of design methodologies relying on techniques like e.g. TRIZ, while the students of the Master of Science in Industrial Design Engineering Technology more rely on general brainstorming techniques or in-house developed

Table 1 Experiment outline (N: no tool; AN: standard AskNature; AN+: standard AskNature with solution principle illustration)

Group A	Group B	Group C	Time	Timeline ↓
Introduction to the experiment				
Explanation AskNature				
Design Problem (DP1)				
N	N	N	15 min.	
Design Problem (DP2)				
N	Strategy list		3 min.	
	AN	AN+	27 min.	
Design Problem (DP3)				
Strategy list		N	3 min.	
AN	AN+		27 min.	
Experiment Debriefing				

creativity stimulation tools. No participants from both studies had been taught any background in Biologically-Inspired Design before the proposed experiment. The two student groups therefore were experienced with systematic design and had a relevant background for this experiment. Participants from Ghent and Leuven University, 59 in total, were divided in three groups, further referred to as A, B and C, respectively containing 20, 20 and 19 students. Random group assignment was approximated by handing participation numbers — according to the pattern A1, B1, C1, A2, B2, C3, etc. — to the students while entering the classroom and by linking these numbers to seats (Verhaegen, 2013).

3.2 *Design cases*

The three design problems were taken from the experimental validation of the PAnDA tool in Design-by-Analogy (Verhaegen, Peeters, Vandevenne, Dewulf, & Duflou, 2011). At the time, the subjects of these design problems were randomly selected from a set of 1011 products from the Google Product Taxonomy (Verhaegen, D'hondt, Vandevenne, Dewulf, & Duflou, 2011). The challenges were: the development of a safer necktie in an industrial context (1), of methods for regulating airflow in a carburetor (2) and of methods for hair removal (3). By repeating these design cases from a validation experiment independent of the current SBID evaluation initiative, more objective case selection was aimed for.

The first design problem required students to develop concepts for a safer necktie. It was explained that accidents can occur in an industrial context when neckties get stuck in machines, resulting in possible injury or even suffocation. The second design problem asked to create concepts for regulating the air-fuel mixture flow in a carburetor. A simplified illustration of a carburetor was shown to the students whose attention was drawn to the valve that currently regulates the flow of the air and fuel mixture. The third design problem challenged the students to develop conceptual solutions for removing body hair, i.e. an alternative for a typical razor.

3.3 *Test conditions and stimuli*

The experiment aims at quantifying the impact of presenting the natural-language biological strategies resulting from a relevant Biomimicry Taxonomy class. Hence the test conditions represent the information returned from problem formulation and search, respectively SBID process steps 1 and 2 (Vandevenne et al., 2014). For AskNature these steps translate to the selection of a relevant functional class and the retrieval of the biological strategies previously assigned to this class by AskNature experts. The relevant functional classes were determined with a small pretest, these are for the Carburettor challenge *get, store, or distribute resources/distribute/gases*, and for the Razor challenge *break down/physically break down/biotic materials*. During the

experiment the participants subjected to the tool conditions received the search results which they needed to sift through to retain the most relevant biological strategies in the offered functional class (SSBID process step 3: filter and analyses) and to formulate analogous technical concepts (SSBID process step 4: knowledge transfer).

Although AskNature is an online, computer-based tool, the experiment outlined in the next section is paper-based to limit the number of factors that potentially could compromise a controlled experiment, such as the use of a computer for other activities than the test, distraction from aspects of the AskNature online tool that are not subject of the test, etc. The results for the design challenges were captured in March 2015. A paper bundle was compiled in March 2015 for each student with the title, short text and photographic image for each strategy in the selected relevant functional class. This stimulus condition is further referred to as the *Standard AN* condition, abbreviated as *AN* (see [Table 1](#)). As it was observed that most images are a visual representation of the involved organism, a second experimental condition was created for which the photographic image is replaced with a graphical illustration of the solution principle, further referred to as the *standard AN+ illustration* condition and abbreviated as *AN+*. For example, for illustrating nest ventilation (prairie dogs) the photographic image of the prairie dog (*AN*) was replaced by a drawing of the tunnel structure which these organisms build in which the induced ventilation flows were indicated (*AN+*). Both conditions were accompanied by the same short natural-language explanation of the strategy, as offered by AskNature.

3.4 Experiment structure and time line

The experiment structure and time line is shown in [Table 1](#). After an introduction to the experiment, a short explanation about AskNature was given explaining how one would identify the biological stimuli like those presented during the upcoming experiment (see [Section 1.1](#)). Before each individual session, the participants were instructed to generate as many different concepts for the problem at hand as possible within the available time frame. No explanation was given about the goal of the research, about the experiment setup or the different experimental conditions.

Next, three design sessions were sequentially executed, see [Section 3.2](#) for the three presented challenges. The first session of 15 min was a personal brainstorm without any aid, hence [Table 1](#) shows the no tool condition (*N*) for groups A, B and C. The output of this first session served to measure the performance of the randomly assigned groups on the adopted metrics (quantity, variety, novelty and quality). If a certain group is at an advantage — i.e. by accidentally containing more productive individuals for a certain metric — this needs to be taken into account during the analyses of the outcomes of the second and third brainstorm to avoid overestimation of the impact caused by the tool conditions.

During the second and third session, one test group was subjected to the control condition (N), one to the standard AskNature condition (AN) and one to the enhanced AskNature condition (AN+). A 3 min table of content analysis was integrated in the non-control conditions to mimic AskNature’s online title-based strategy exploration which is presented to the online users after selection of a functional class that matched their design challenge. At the time of writing, this functionality is offered via the ‘explore by function’ link on AskNature’s home page. Next, the participants in the tool conditions were instructed to solve the challenge using the provided stimuli if these were deemed useful. As the second and third brainstorm involved processing of natural-language stimuli from a different domain, a total of 30 min were allocated per challenge. Condition-to-group assignment was alternated between the second and third session to allow each participant the experience of using at least one AskNature condition variant. For both sessions, the control condition served as baseline to which the performance of the AN and AN+ groups was compared.

4 Results

The results of the above outlined experiment are detailed in the four subsections below, one for each metric. In these subsections the effect of the test conditions (AN and AN+) is measured against the control condition (N) for the second and third brainstorm. The first brainstorm serves as verification of the test group balance, i.e. a potential source of covariates. Before discussing the results, the global approach for statistical significance estimation is outlined in the next paragraphs.

In order to quantify the impact of the tool conditions on the four metrics, the mean score of the metric per test condition provides a first indication, e.g. column 2 of [Table 2](#) demonstrates a mean quantity score for both tool conditions that is considerably lower compared to the no tool condition. Next, the hypothesis of equal group means is tested with an analysis of variance (AN-OVA), a generalization of the t-test for more than two groups. Hence, the null hypothesis states that all conditions (N, AN, AN+) have the same or no effect. A low p -value — alpha is set to 0.05 for all analyses — justifies the rejection of this null hypothesis, and thus indicates a statistically significant effect for the corresponding condition. See, for example, columns 3 and 4 of [Table 2](#), respectively indicating the effect and the significance hereof.

Although random participant-to-group assignment aims for similarly composed groups — potential sources of variability on the metrics’ results that are not of primary interest are investigated. First, for each metric and each brainstorm, two potential fixed factors — educational background and gender — are evaluated with Pearson’s correlation test: r indicates the strength of the correlation (between -1 and 1) and p indicates the two-tailed

Table 2 ANCOVA for the quantity metric for the Carburettor and Razor challenges^a

	<i>Carburettor</i>			<i>Razor</i>		
	<i>Score</i>	<i>Effect</i>	<i>p-value</i>	<i>Score</i>	<i>Effect</i>	<i>p-value</i>
N	7.736	N/A	N/A	5.842	N/A	N/A
AN	4.941	-36.13%	0.001	4.684	-19.82%	0.122
AN+	4.526	-41.49%	<0.001	5.611	-3.95%	0.758

^a See Section 4.1 for fixed factors and covariates.

significance of the result, i.e. the probability of detecting a correlation while the variables are not correlated. Next, potential covariates are identified by evaluating the difference between the groups for each metric in the first brainstorm (i.e. potential violation of group balance) and by performing a correlation test with the metric's results of the first brainstorm to the other brainstorms.

Depending on the identified fixed factors and covariates, the most appropriate statistical model for the results below might be a standard analysis of variance (ANOVA) or an analysis of covariance (ANCOVA). Furthermore, when one of the assumptions required for the validity of the statistical model fails, appropriate actions are taken, e.g. the use of a Welch test instead of the standard ANOVA test when homogeneity of variance is violated. To ensure readability of the next subsections, only the statistical results with noteworthy impact to the hypotheses are detailed. Lower level statistical details are omitted such as, for example, all the corresponding r and p -values of the fixed factors or covariates that are not retained. The most important results are summarized in tabular format (Tables 2–5), where the score column represents the mean value for the specific metric for each condition, the effect column shows the tool conditions' relative effect compared to the no tool condition and p -value column reflects the significance.

4.1 Quantity

The participants' background has a weak but statistical significant effect on their total number of generated concepts in the third brainstorm ($r = 0.308$, $p = 0.02085$). Hence background is taken as a fixed factor for quantity of the third brainstorm. As for the participants' background, their gender also had a relatively weak but significant effect on the number of concepts they generated during the third challenge ($r = 0.295$, $p = 0.021$). The mean number of concepts for females was 6.77 and for males 5.04. This observation indicates that female participants were more fluent in generating different ideas for body hair removal. Hence, gender is also taken as a fixed factor for this third challenge. No significant difference between the three groups in the baseline test (i.e. the first design case) was measured for quantity (ANOVA with $p = 0.535$), but the number of concepts generated by the participants during this first brainstorm — where all groups were subjected to the no tool

Table 3 Results of the level-based variety metric for the Carburettor and the Razor brainstorm^a

		<i>Carburettor</i>			<i>Razor</i>		
		<i>Score</i>	<i>Effect</i>	<i>p-value</i>	<i>Score</i>	<i>Effect</i>	<i>p-value</i>
Physical principle variety	N	1.973	N/A	N/A	5.527	N/A	N/A
	AN	3.710	+88.04%	0.009	5.110	-7.54%	0.8568
	AN+	4.108	+108.21%	0.001	5.181	-6.25%	0.8667
Working principle variety	N	4.565	N/A	N/A	7.713	N/A	N/A
	AN	7.056	+54.58%	0.005	7.633	-1.04%	0.990
	AN+	7.122	+56.03%	0.003	7.812	+1.28%	0.986

^a See Vandevenne, Pieters, et al. (2015) for more details on the statistical analyses.

Table 4 Results of the novelty metric for the Carburettor and the Razor brainstorm^a

	<i>Carburettor</i>			<i>Razor</i>		
	<i>Score</i>	<i>Effect</i>	<i>p-value</i>	<i>Score</i>	<i>Effect</i>	<i>p-value</i>
N	2.650	N/A	N/A	2.835	N/A	N/A
AN	3.866	+46.90%	0.055	4.041	+42.55%	0.001
AN+	4.142	+56.31%	0.017	4.736	+67.07%	<001

^a Standard ANOVA for the Carburettor and Welch test for the Razor brainstorm.

Table 5 Quality (technical feasibility) for the Carburettor and Razor brainstorm^a

	<i>Carburettor</i>			<i>Razor</i>		
	<i>Score</i>	<i>Effect</i>	<i>p-value</i>	<i>Score</i>	<i>Effect</i>	<i>p-value</i>
N	1.970	N/A	N/A	1.906	N/A	N/A
AN	1.877	-4.72%	0.106	1.918	+0.63%	0.952
AN+	1.948	-1.12%	0.811	1.919	+0.68%	0.917

^a Welch test for the Carburettor and standard ANOVA for the Razor brainstorm.

condition — correlates to their number of concepts generated in the second ($r = 0.472$, $p < 0.001$) and third brainstorm ($r = 0.385$, $p = 0.003$). Hence, this variable is taken as covariate for the analyses of the conditions' effects in the second and third challenge to account for individual's fluency in concept generation.

The results for the Carburettor and Razor challenge are shown in Table 2 and illustrated by Figure 2. For the Carburettor brainstorm, analysis of covariance shows statistically significant difference ($p < 0.001$) between the means of the three test groups, hence the test conditions (N, AN and AN+) have a significant influence on the number of generated concepts. Post hoc analysis reveals a significant difference between the N and AN condition ($p = 0.001$) and between the N and AN+ condition ($p < 0.001$); and descriptive statistics show that the tool conditions (AN and AN+) cause a decrease in the number of

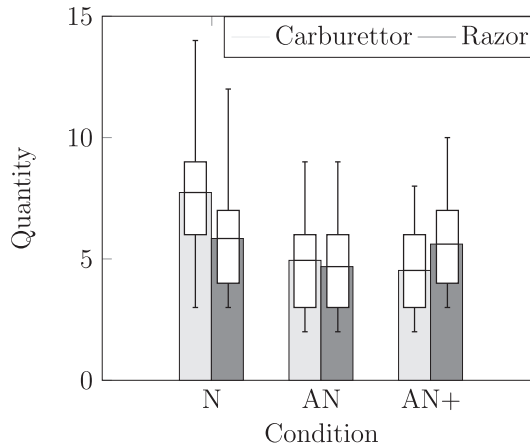


Figure 2 Quantity results for the Carburettor and Razor session

generated concepts. This is visualized by Figure 2 by the Carburettor boxplots and by the mean score column in Table 2.

For the third brainstorm, there is no significant difference between the three test groups ($p = 0.231$), although the mean quantity score also decreases for both tool conditions compared to the no tool condition. The AN condition demonstrates a 19.82%, but not significant, lower mean quantity score compared to the N condition. The group subjected to the AN+ condition almost demonstrates a status quo. This lower drop in quantity could be caused by the fact that the participants who received the AN+ condition in the third challenge were more experienced as they were also subjected to the AN condition in the second challenge. This experience might have helped the participants in more efficiently sifting through the biological strategies, by e.g. utilising the table of contents more efficiently.

The decrease in number of generated concepts for the tool conditions could be caused by the effort required to use the AN and AN+ conditions. For these conditions, the table of contents is first browsed and the strategies deemed relevant are analysed in detail. This requires a considerable amount of time per candidate biological solution, while the participants without a tool (N) generate concepts based on their personal experience and knowledge.

4.2 Variety

Level-based variety analysis for the above experimental setup has been described and illustrated by Vandevenne, Pieters, Vanneste, and Duflou (2015). The results, which were found to be case-dependent, are summarised in Table 3. Indeed, for the Carburettor brainstorm the average variety score increased significantly for both tool conditions (AN and AN+) on both the physical and working principle level of the genealogy tree. This indicates

that, for this second brainstorm, the tool assisted in generating a more diverse idea set. However, the results from the Razor brainstorm did not confirm these findings.

An important note to make here addresses the impact of quantity on the level-based variety results. As noted by [Verhaegen \(2013\)](#) it is more likely that similar concepts exist in a large set of concepts compared to a small set when adding an additional concept. This means that tools that decrease quantity are at an advantage for variety metric calculation and it is more likely that a more diverse set of concepts (as defined by the variety metric) is generated in such a context. This reasoning is supported by the negative correlations between quantity and physical and working principle variety for each of the brainstorms (r ranges from -0.196 to -0.767) and by the observation that the AN and AN+ conditions that triggered significantly less concepts in the second brainstorm ([Table 2](#)) demonstrated significantly higher level-based variety for this brainstorm ([Table 3](#)).

4.3 Novelty

Novelty is calculated by measuring the rarity of the concepts, with the set of concepts from the non-primed control condition as Universe of Ideas for Comparison. See [Section 1.2.3](#) for the equations used for novelty calculation. No fixed factors (background and gender) or covariates (novelty in the first design challenge) were identified for the novelty metric.

Both the Carburettor and Razor brainstorms demonstrate a large increase in mean novelty score — ranging from $+42.55\%$ to $+67.07\%$ — for both tool conditions (AN and AN+) compared to the no tool condition (N) (see [Figure 3](#)). [Table 4](#) indicates the significance levels for each effect. Novelty increase is statistically supported for all tool conditions in both brainstorms, except for the AN condition during the Carburettor brainstorm where near significance is measured at $\alpha = 0.05$ (p -value = 0.055). These results indicate that using a relevant functional category of AskNature’s Biomimicry Taxonomy has the potential of having a strong positive effect on the average novelty scores of the participants. Furthermore, both average novelty scores for the Carburettor and Razor brainstorm increase more for the AN+ condition than for the AN condition (respectively an extra $+10.41\%$ and $+24.52\%$) and statistical significance of the results strengthens (see [Table 4](#)).

4.4 Quality

No fixed factors (background and gender) or covariates (quality in the first design challenge) were identified for the quality metric. As technical feasibility scores are personal opinions, such scoring is subjective and its repeatability needs verification. Hereto, two judges (the first and second author of this

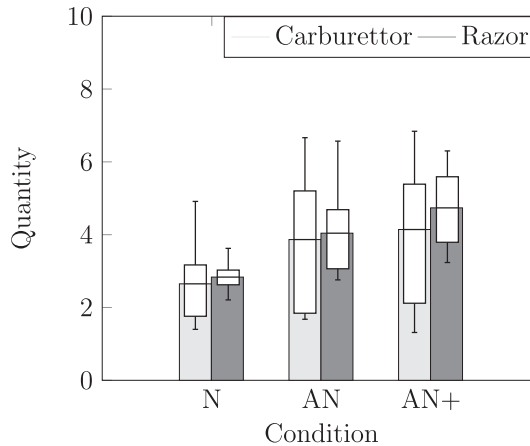


Figure 3 Novelty results for the Carburettor and Razor session

paper) independently applied the scoring mechanism explained in Section 1.2.4 to each unique concept generated during the first brainstorm. For these scores, the free-marginal interrater kappa was calculated (Randolph, 2005). This kappa can range from -1 (perfect disagreement) to 1 (perfect agreement) and takes into account agreement by chance. The obtained value is 0.86 , hence the adopted technical feasibility scoring mechanism can be regarded as relatively consistent between judges.

Table 5 summarises the results for the quality metric. One can see that there is no significant effect observed for any of the test conditions on the average technical feasibility of the produced concepts per participant in the Carburettor and Razor brainstorms. This is, on first sight, a remarkable observation. As Biologically-Inspired Design is expected to support finding out-of-the-box ideas, a derived expectation could be that such novel ideas would be less technically feasible than those from a no tool condition. As the results of Table 5 show no significant effect on the group means, correlation tests are performed to further investigate this anticipated effect. A significant weak negative correlation ($r = -0.343$, $p = 0.010$) between average technical feasibility (quality) and average novelty per participant for the Carburettor brainstorm is identified, but an almost significant weak positive correlation for the Razor brainstorm ($r = 0.238$, $p = 0.077$). Hence, the relation between novelty and quality seems to be not outspoken and case-dependent.

5 Discussion

As AskNature pages typically contain – besides a title and photographic image – a short extract from a biological source, such as a biological book or academic paper, the presentation format of the AN condition can be generalized as a short – i.e. 425 words on average – domain-specific explanation of the

biological solution principle in natural-language format accompanied by a photographic image of the biological system. Hence, similar performance is expected for other tools that present biological excerpts as stimuli in response to a functional query. However, outcome-based performance should be evaluated for each tool individually.

The results show a consistent novelty increase for both the AN and AN+ condition, compared to the control condition, for the Carburettor as well as the Razor brainstorm. The AN conditions added +46.90% in Brainstorm 2 (almost significant with $p = 0.055$) and +42.55% novelty in Brainstorm 3 (strongly significant with $p = 0.001$). Adding the graphical illustrations to the standard AN conditions, results in +56.31% and +67.07% novelty ($p = 0.017$ and $p < 0.001$) for Brainstorms 2 and 3 respectively. The results support the first hypothesis, a relevant functional category from AskNature increases novelty of the design outcomes; and the second hypothesis as the AN+ condition generated higher average novelty with stronger significance. Interestingly, novelty increase triggered by the AN and AN+ conditions did not come at a measured cost to the technical feasibility of the concepts, as no significant effect of the tool conditions was measured on the participants' average quality scores and correlation between novelty and technical feasibility is found to be weak and not consistent.

The results for quantity and variety should be evaluated jointly. For each tool condition the number of concepts decreased and strong significance hereof was measured for the Carburettor brainstorm. As cognitively processing the AN and AN+ stimuli requires considerable time and effort, no drop in quantity would have been surprising within a limited time frame. Stimulating towards fewer concepts should not necessarily be judged negatively, as it is exactly the goal of the supporting tool to focus users on relevant suggestions. It is better to obtain a low number of highly novel and technically feasible concepts, than a large number of concepts with less favourable characteristics. As it is inherent to the variety metric that higher variety is more easily obtained for smaller idea sets, it is logical that variety results inversely follow the quantity results, and significance was also measured only for the Carburettor brainstorm (both physical and working principle variety).

The current test setup measures the potential of using one relevant functional class from AskNature's Biomimicry Taxonomy for two challenges that apply to randomly chosen products. For the Carburettor problem the *get, store, or distribute resources/distribute/gases* class held 23 biological strategies, for the Razor challenge the *break down/physically break down/biotic materials* offered 17 biological strategies. It is important to note that currently AskNature's knowledgebase does not hold a sizeable set of biological solutions for each functional class. On average there are 19 strategies classified per function, with some categories containing over 50 and some categories containing

zero strategies. It is evident that an extremely low number of biological solutions in the relevant class would heavily impact performance, hence *potential* performance for well-supported Biomimicry Taxonomy classes is reported in this paper. Furthermore, it is advisable for future work to draw new test challenges that apply for different functional classes to further confirm the above findings.

6 Conclusions

This contribution presents the first quantitative assessment — based on the four commonly used validation metrics quantity, quality, variety and novelty — of a publicly accessible knowledge-based bio-ideation tool. AskNature was chosen for this purpose as it is by far the most popular and well-known BID tool. The effect of two conditions — standard AskNature format (AN) and standard AskNature format in which the photographic image is replaced by a graphical illustration of the solution principle (AN+) — is measured and compared to a no tool control condition. Hereto, the output of the experiment conducted with 59 participants — generating 945 concepts — was analysed in an outcome-based study. The four typical outcome-based metrics were used: quantity, variety, novelty and quality. The first two, quantity and variety representing respectively the number of concepts and the diversity of ideas on different abstraction levels, are considered as means to an end. The last two, novelty and quality, measure more directly the ideation goal of identifying out-of-the-box, still technically feasible, concepts.

In terms of metrics, a popular argument in favour of Biologically-Inspired Design is its potential for finding out-of-the-box solutions. Indeed, the experiment consistently found large novelty increases ranging from +42.55% to +67.07% depending on the specific AskNature condition and brainstorm. The proposed AN+ condition resulted in higher novelty scores with stronger statistical significance compared to the AN condition, indicating a potential merit of the new proposed stimulus condition. Based on the presented results, it is concluded that presenting AskNature's functional classes holds strong potential for identifying out-of-the-box solutions during ideation and that replacing the photographic illustration of the organism by a graphical representation of the solution principle further strengthens this potential.

Furthermore, the results indicate that using a biological stimulus representation in the format of a title, image, and short biological excerpt explaining the solution principle is likely to have a dampening impact on the number of generated ideas, compared to a no tool control condition. This illustrates the cognitive effort required to explore the presented stimuli, to process and understand the biological solution principle in its current format and to formulate an analogous technical concept. Hence, the tests indicate that this

priming ideation tool steers towards fewer concepts in the same amount of time (30 min), with higher novelty and comparable technical feasibility.

Future work should focus on further validating the current results with more and different challenges, on further experimenting with new stimuli representations to facilitate analyses and cross-domain knowledge transfer, on measuring the consistency of identifying relevant Biomimicry Taxonomy classes and on validating other bio-ideation approaches and tools.

References

- Bar-Cohen, Y. (2011). *Biomimetics: Nature-based Innovation*. Boca Raton, FL: CRC/Taylor & Francis.
- Bonser, R. H. (2006). Patented biologically-inspired technological innovations: A twenty year review. *Journal of Bionic Engineering*, 3, 39–41.
- Chakrabarti, A., Sarkar, P., Leelavathamma, B., & Nataraju, B. (2005). A functional representation for aiding biomimetic and artificial inspiration of new ideas. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 19, 113–132.
- Dahl, W. D., & Moreau, P. (2002). The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research*, 39, 47–60.
- Dean, D., Hender, J., Rodgers, T., & Santanen, E. (2006). Identifying quality, novel, and creative ideas: Constructs and scales for idea evaluation. *Journal of the Association for Information Systems*, 7, 646–698.
- Deldin, J., & Schuknecht, M. (2014). *The AskNature Database: Enabling Solutions in Biomimetic Design*. Chapter 2 in *Biologically Inspired Design*. London: Springer-Verlag 17–27.
- Durand, F., Helms, M., Tsenn, J., McTigue, E., McAdams, D., & Linsey, J. (2015). Teaching students to innovate: Evaluating methods for bioinspired design and their impact on design self efficacy. In *27th International Conference on Design Theory and Methodology*, 7.
- Fish, F., & Beneski, J. (2014). Evolution and bio-inspired design: Natural limitations. *Biologically Inspired Design* 287–312.
- Gebeshuber, I., Gruber, P., & Drack, M. (2009). A gaze into the crystal ball: Biomimetics in the year 2059. *Journal of Mechanical Engineering Science*, 223, 2899–2918.
- Glier, M., Tsenn, J. J., & McAdams, D. (2012). Evaluating the directed method for bioinspired design. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 403–413).
- Herfindahl, O. (1950). *Concentration in the Steel Industry*. Ph.D. thesis. Columbia University.
- Kaiser, M., Farzaneh, H., & Lindemann, U. (2014). Bioscrabble the role of different types of search terms when searching for biological inspiration in biological research articles. In *Proc. of Design 2014* (pp. 241–250).
- Kim, J., McAdams, D., & Linsey, J. (2014). Helping students to find biological inspiration: Impact of valuableness and presentation format. In *Frontiers in Education Conference (FIE)* (pp. 1–6).
- Lenau, T., Dentel, A., Ingvarsdóttir, T., & Gudlaugsson, T. (2010). In *Engineering Design of an Adaptive Leg Prosthesis using Biological Principles* (pp. 331–340).

- Linsey, J. S. (2007). *Design-by-Analogy and Representation in Innovative Engineering Concept Generation*. Ph.D. thesis. Austin, TX: The University of Texas.
- Nagel, J., & Stone, R. (2012). A computational approach to biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 26, 161–176.
- Nelson, B., Wilson, J., & Yen, J. (2009). A study of biologically-inspired design as a context for enhancing student innovation. In *Frontiers in Education Conference, 2009. FIE'09. 39th IEEE* (pp. 1–5).
- Randolph, J. J. (2005). *Free-marginal Multirater Kappa: An Alternative to Fleiss' Fixed-marginal Multirater Kappa*. Joensuu, Finland: Joensuu University Learning and Instruction Symposium.
- Schild, K., Herstatt, C., & Lüthje, C. (2004). *How to Use Analogies for Break-through Innovations*. Institut Technologie und Innovationsmanagement. Working Paper; No 25.
- Shah, J. J., Kulkarni, S. V., & Vargas-Hernandez, N. (2000). Evaluation of idea generation methods for conceptual design: Effectiveness metrics and design of experiments. *Journal of Mechanical Design*, 122, 377–384.
- Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, 24, 111–134.
- Shu, L. (2010). A natural-language approach to biomimetic design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 24, 507–519.
- Tseng, I., Moss, J., Cagan, J., & Kotovsky, K. (2008). The role of timing and analogical similarity in the stimulation of idea generation in design. *Design Studies*, 29, 203–221.
- Vandevenne, D., Pieters, T., Vanneste, C., & Dufloy, J. (2015). Case-dependence of level-based idea space variety for systematic biologically-inspired design. In *Proceedings of TRIZ Future Conference*. (p. Status: Accepted for publication).
- Vandevenne, D., Verhaegen, P.-A., Dewulf, S., & Dufloy, J. (2011). A scalable approach for the integration of large knowledge repositories in the biologically-inspired design process. In *Proc. of ICED 2011*.
- Vandevenne, D., Verhaegen, P.-A., Dewulf, S., & Dufloy, J. (2014). A scalable approach for ideation in biologically inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 29, 19–31.
- Vandevenne, D., Verhaegen, P.-A., & Dufloy, J. (2015). SEABIRD: Scalable search for systematic biologically-inspired design. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 30, 78–95.
- Vattam, S., Wiltgen, B., Helms, M., Goel, A., & Yen, J. (2010). DANE: Fostering creativity in and through biologically inspired design. In *Proc. of First International Conference on Design Creativity, Kobe, Japan*.
- Verhaegen, P.-A. (2013). *Methods and Algorithms for Systematic Innovation*. Ph.D. thesis. KU Leuven.
- Verhaegen, P.-A., D'hondt, J., Vandevenne, D., Dewulf, S., & Dufloy, J. (2011). Identifying candidates for design-by-analogy. *Computers in Industry*, 62, 446–459.
- Verhaegen, P.-A., Peeters, J., Vandevenne, D., Dewulf, S., & Dufloy, J. (2011). Effectiveness of the panda ideation tool. *Procedia Engineering*, 9, 63–76.
- Verhaegen, P.-A., Vandevenne, D., Peeters, J., & Dufloy, J. R. (2013). Refinements to the variety metric for idea evaluation. *Design Studies*, 34, 243–263.
- Vincent, J., Bogatyreva, O., Bogatyrev, N., Bowyer, A., & Pahl, A. (2006). Biomimetics: Its practice and theory. *Journal of the Royal Society: Interface*, 3, 471–482.

- Ward, T. B. (1998). Analogical distance and purpose in creative thought: Mental leaps versus mental hops. In *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences* (pp. 221–230).
- Wilson, J., Rosen, D., Nelson, B., & Yen, J. (2010). The effects of biological examples in idea generation. *Design Studies*, 31, 169–186.