Proceedings of the 2021 ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2021
August 16-19, 2021, Virtual

DETC2021-71905

ALIGNING HUMAN AND COMPUTATIONAL EVALUATIONS OF FUNCTIONAL DESIGN SIMILARITY

Ananya Nandy
Dept. of Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
ananyan@berkeley.edu

Kosa Goucher-Lambert*
Dept. of Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
kosa@berkeley.edu

ABSTRACT

Function drives many early design considerations in product development. Therefore, finding functionally similar examples is important when searching for sources of inspiration or evaluating designs against existing technology. However, it is difficult to capture what people consider to be functionally similar and therefore, if measures that compare function directly from the products themselves are meaningful. In this work, we compare human evaluations of similarity to computationally determined values, shedding light on how quantitative measures align with human perceptions of functional similarity. Human perception of functional similarity is considered at two levels of abstraction: (1) the high-level purpose of a product, and (2) a detailed view of how the product works. Human evaluations of similarity are quantified by crowdsourcing 1360 triplet ratings at each functional abstraction, and then compared to similarity that is computed between functional models. We demonstrate how different levels of abstraction and the fuzzy line between what is considered "similar" and "similar enough" may impact how these similarity measures are utilized, finding that different measures better align with human evaluations along each dimension. The results inform how product similarity can be leveraged by designers. Therefore, applications lie in creativity support tools, such as those used for design-by-analogy, or future computational methods in design that incorporate product function in addition to form.

1 Introduction

Designers often make comparisons between different ideas and assess how their designs will meet functional requirements to solve the problems they face. In order to acquire knowledge and make decisions in early-stage design, a common practice is to seek examples of or inspiration from existing products, through methods such as benchmarking or searching patents [1,2]. Previous work has shown that inspirational stimuli help improve idea generation and that function-based examples are specifically useful in helping designers identify potential solutions [3,4]. Therefore, many quantitative approaches have been applied to determine functional similarity between products, guiding the development of computational methods to augment designers' capabilities in the solution exploration phase [5–7]. However, it is still challenging to determine how these methods align with the way designers draw functional connections between products in practice. Depending on the stage of the design process, designers may consider concepts at different levels of abstraction. While functional representations are abstractions themselves, prioritization may lie on the higher-level function of the product (i.e. its purpose) or on the mechanisms/sub-functions necessary to achieve this purpose (i.e. how it works) [8]. When considering function in these different ways, products that are relevant for benchmarking, as examples, or for inspiration, may vary, motivating the need for appropriate similarity measures.

To further assess how design similarity can be measured and utilized for the early stages of design, in this research we take a quantitative approach to compare human conceptualization of

*Address all correspondence to this author.

similarity with how similarity can be measured mathematically, specifically focusing on product function. We apply crowdsourcing methods to quantify human-determined functional similarity and explore how various quantitative measures align with human understanding of functional similarity. The results lead to insight on how measures of functional similarity can be used in engineering design and where they might fall short. With our contributions, we aim to ensure that computational methods to support design are meaningful to humans.

2 Related Work

In this section, we review prior research on the use of similarity measures in engineering design, as well as research on human perception of similarity and its measurement. These areas are applicable to our work in comparing human evaluations with computational output, ensuring that the latter is interpretable and useful to humans during design activities.

2.1 Measuring similarity in design

Obtaining evaluations of design similarity from humans, through expert or crowdsourced assessments, is challenging and expensive, prompting effort towards finding quantitative similarity measures for the design domain. Since design spans various tasks and contexts, it is often practical and desirable to adapt existing measures that have already been shown to apply across various domains. Similarity can be assessed on different dimensions, such as form or function, and at different phases, ranging from concepts to full products [9]. For example, visual similarity between products (similarity in form) has been investigated for the purpose of determining product families, variants, or branding [10,11]. In the early stages of design, however, product function is often one of the most critical considerations [2]. Assessing similarities along the dimension of product function poses a challenge because product function is difficult to quantify.

One way to calculate functional similarity involves using a text-based repository of domain relevant information such as the patent database, which contains a large body of data on product function. Functional similarity has been calculated using latent semantic analysis and the cosine similarity measure on these patents for the purpose of design-by-analogy (a method where designers seek to apply solutions that work for other problems to solve their problem) [5]. The results from using this measure has been validated by indicating that its clustering of patents is sensible to experts [12]. Another way to capture product function is through a functional diagram or model, such as one developed using a standardized vocabulary [2,13,14]. A vector-based quantitative metric has been developed to compare these functional models [6]. In addition, critical function chains have been extracted from functional models and matched in various ways to quantify functional similarity [7, 15]. The functional model representation enables a higher level of abstraction of a product than a patent, which may be desirable when searching for examples during conceptual design. At the same time, functional models are not available for many products and are often developed subjectively. To mitigate these challenges, recent research has focused on automating functional modeling using information about product components [16, 17]. However, it is also notable that these measures each provide a different conceptualization of what similarity means when applied to product function. Some measures place importance on the existence (or absence) of specific sub-functions to define overall functional similarity, while others place more importance on patterns in how sub-functions connect to each other. These differences impact the context each measure can be used within and indicates the necessity to carefully consider how functional similarity is quantified [18].

2.2 Considerations for evaluating design similarity

Several factors may influence how quantitative similarity measures, computed directly on designs, can be effectively used to support human creativity in engineering design. Two are considered here: first, that there is no hard line between when a product is considered similar vs. dissimilar, and second, that ideas are considered at varying abstraction levels during design.

The concepts of similarity, dissimilarity, and distance are often used interchangeably. Mathematically, distance measures can be converted to dissimilarity measures. In addition, similarity can be converted to dissimilarity, and vice versa. However, in application, whether similarity or dissimilarity is more important depends highly on the context. For example, in recommendation systems broadly, while similarity is used to find the most relevant results, the notion of dissimilarity is explored instead to add novelty and diversity to results [19]. In the domain of engineering design, dissimilarity has been applied to developing novelty metrics for idea assessment at the conceptual stage. To characterize novelty, these metrics emphasize how dissimilar a given concept is from other concepts [20].

Measures of similarity also play a critical role in attempts to foster analogical innovation. According to Gentner and Markman, the processes involved in comparisons of similarity and analogy are the same [21]. Analogical distance has been shown to impact the effectiveness of examples during concept generation, indicating that the distinction between similarity and dissimilarity is critical in the practice of design-by-analogy [22,23]. Analogies are generally considered to be near-field analogs (sharing surface features or existing in the same domain) or far-field analogs (sharing few or no surface features and existing in different domains, but having some functional similarity). As such, analogical distance encompasses similarity and dissimilarity as well as balancing the line between the two [24]. It would thus be essential to understand if the suitability of a measure to find similar designs differs from the suitability of the measure to find dissimilar designs.

Another element to consider when assessing similarity between designs or products is the abstraction level of their representations. Work in the cognitive processes behind design suggests that solution search is performed through lateral and vertical transformation: moving to a slightly different idea or moving to a more detailed version of the same idea [25]. In the context of product function, as the level of detail available increases, the functional abstraction can decrease, facilitating consideration of function to be how the product works instead of its higher-level purpose [26]. Functional similarity measures have focused on lower-level representations (i.e. the working mechanism) since very detailed information is available in patents and in full functional models that have been developed through reverse engineering. However, designers often only have enough information to operate at the higher level during the conceptual stages of design. In addition, cross-domain analogies can be found through higher-level purpose even if working mechanisms differ [8]. To address that functional similarity can and should be considered at different levels of abstraction, it is necessary to understand how any quantitative measures reflect the ways in which humans translate between the levels.

2.3 Quantifying how humans evaluate similarity

Humans constantly make judgements of similarity in order to reconcile information from the world around them with internal mental representations. In addition, similarity is said to play a part in how people structure conceptual knowledge [27]. In the structural alignment view of similarity from psychology, there are three elements of alignment: structural consistency, relational focus, and systematicity. These elements correspond to one-to-one matching, common relations in both items being compared, and sets of relations that are interconnected by higher order relations [21]. It is difficult to untangle the underlying dimensions along which people consider similarity, which is important when applying similarity measures within interventions or systems that are intended to augment human processes. In order to assess how quantitative measures align with human mental models, it is crucial to capture how humans perceive similarity.

One approach to understanding why people might consider two objects to be similar is to explicitly ask them for their reasoning. Another is a data-driven approach, where people are asked to make similarity judgements and latent or explainable dimensions are uncovered directly from the results. This data-driven approach has been used in several contexts such as to determine similarity across musical artists and natural objects [27,28]. Within engineering design, both approaches have been used to assess design similarity for a variety of purposes.

In a study on design-by-analogy, participants were explicitly asked what dimensions they considered important for similarity between a target and source product [29]. It was found that functional similarity dominated over form similarity. To understand whether the structural alignment view of similarity from psychology applies in the context of design, participants were asked to rate similarity between design concepts and explain their reasoning in another study [30]. The results implied that feature-based responses drove similarity, in line with the element of structural consistency from the structural alignment model.

More recently, the data-driven approach has been increasingly applied to problems in design. Pairwise similarity judgements were crowdsourced to assess visual similarity between products, determining that novelty assessments from a crowd can match with those made by experts [31]. Similarity judgements were also collected in the form of triplet ratings for determining design sketch novelty and for evaluating dissimilarity between sets of ideas to spur diversity during idea generation [32, 33]. More granular search across specifically product function was enabled by crowdsourcing annotations from product descriptions and including both product purpose and working mechanism as facets [34]. These data-driven methods are able to uncover human perceptions of similarity, but may be limited to the task or context for which the data were collected. In addition, the dimensions of similarity determined from data-driven methods may not be explanatory or easy to interpret. Therefore, using similarity functions that have been learned directly from humans may not always be possible or desirable. At the same time, even a similarity measure that is directly computed from products must provide human-interpretable results in order to successfully supplement cognitive processes such as analogical transfer.

3 Methods

In this paper, we investigate functional similarity as crowdsourced from humans in comparison with what is recovered using quantitative measures directly on the products. We aim to gain insight into how example-based design tools or computational methods might capture product functionality, making this information accessible to designers.

3.1 Product dataset

A subset of 20 consumer products (e.g. toys, consumer electronics, household devices) found in the Design Repository hosted by Oregon State University was utilized for this work [35]. This subset was selected to represent products with varying levels of complexity that participants would be familiar with, as well as to ensure the availability of two consistent levels of functional specification. A list of the products can be found in Appendix A. For each product, the repository contained a simple functional model consisting of inputs, outputs, and a singular, main function of the product, as well as a highly detailed functional model of how the product worked, specified according to Hirtz et al. [14]. An example of each type of functional model is shown in Figure 1. The repository additionally contained a product title and image.

3.2 Crowdsourcing human judgements

To capture how humans consider products to be functionally similar, similarity judgements were crowdsourced from humans and then represented in a two-dimensional embedding space us-

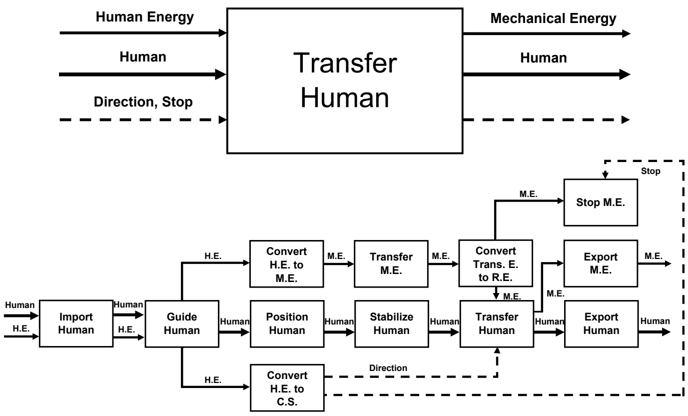


FIGURE 1: Functional models at two levels of abstraction (shown for a scooter)

FIGURE 2: Example of a triplet query. The function description text displayed under the three images were descriptions from one of the two levels of abstraction (detailed in Appendix A. Here, the higher abstraction is shown.

ing techniques from machine learning. These embeddings were used to quantify the relative similarity among the set of products. This method has recently been used in engineering design to determine the visual similarity of products as well as to determine the novelty of ideas [31, 32]. The judgements were collected in the form of triplet queries ("Is A more similar to B or to C?"). Prior work has shown that humans are more easily and consistently able to answer triplets as opposed to direct pairwise comparisons [36].

Because functional similarity may depend on the level of abstraction at which someone is considering products, information about function was presented to participants in two ways based on the two available types of functional models. The information from the full functional models (lower image in Figure 1) was converted to text descriptions to capture essential information about how the product worked. The function as defined by its purpose was taken directly as text from the simple functional model (upper image in Figure 1) and only modified in a few cases for clarity. The descriptions can be found in Appendix A. Stock images were included for products that were missing images in the Design Repository and product titles were modified to represent the generic versions of the product. Each triplet presented to participants contained the following information about each product: a title, image, and description of the product function of either type. An example triplet is shown in Figure 2.

Although images were provided to aid them in understanding what the products were, participants were instructed to judge similarity along the dimension of function and not form. The participants were instructed to consider the overall purpose of the product when presented with the shorter descriptions and to consider the way in which the product worked when presented with the longer description. Each participant made judgements on the same subset of triplets twice, once presented with the shorter descriptions and once presented with the longer description.

After approval from an Institutional Review Board, data was collected from a total of 69 participants. Data from one participant was removed, as they did not follow instructions. The included participants consisted of 42 undergraduate students, 16 graduate students, and 10 others (including working professionals), and among them were 50 participants who identified as male, 17 who identified as female, and 1 who preferred not to say. A majority of participants were pursuing, or had graduated with, a mechanical engineering degree. 24 of the participants indicated that they had greater than 4 years of engineering/design experience through courses, work, or extracurricular activities. 36 of the participants were shown the longer, lower-abstraction descriptions first, while 32 were shown the shorter, higher-abstraction descriptions first.

A total of 2720 triplet ratings were collected from the participants, who each provided 40 ratings. Half of each set of ratings (1360 triplets) contained information about each level of functional abstraction. Therefore, each set of triplets collected consisted of 42 percent of the total possible triplets (3240 triplets). Previous applications of the two-dimensional embedding techniques considered here use 20 percent of triplets [37, 38]. Additionally, prior work that incorporates the full triplet set found that using 30 percent of the triplet set was sufficient and that the methods were robust to a small number of false ratings [32].

3.3 Generating a 2D embedding space

Once the triplets were collected, they were mapped to a 2D space. The two-dimensional data embeddings can be constructed by using the triplets as constraints to where points (in this case, products) are placed within the 2D space. One method to do this is t-Distributed Stochastic Triplet Embedding (t-STE). This method defines a probability density distribution (a heavy-tailed kernel) and maximizes these probabilities with respect to the embedding points so that a triplet is satisfied. Additionally, the maximization ensures that similar points are collapsed while dissimilar points kept apart by triplets are repelled [38]. Other commonly used embedding methods include Generalized

Non-metric Multidimensional Scaling (GNMDS), Crowd Kernel Learning (CKL), and Stochastic Triplet Embedding (STE) [36–38]. GNMDS and t-STE have been explicitly applied to work in the design domain [32,33]. In this work, t-STE was chosen as the embedding method, due to its ability to ensure similar points are closer together and dissimilar points are farther apart, without violating any constraints [38]. Preliminary analyses demonstrated that selecting t-STE as the embedding method (as opposed to one of the aforementioned techniques) did not significantly affect the results.

In addition, to address partial triplet collection and the aggregation of triplet ratings from across the population of participants, two measures, adapted from prior work, were used to determine the quality of the embedding: distance error and triplet generalization error. Distance error refers to the mean squared error between the the normalized Euclidean distances derived from the final embedding and an embedding created with consecutively fewer triplets [32]. This measure was used to determine how much the embedding changes with the addition of new triplets in order to ensure that there are enough triplets for convergence. Triplet generalization error is calculated by holding out a set of triplets, calculating the embedding, and then determining whether the calculated embedding satisfies the triplets that were held out [38]. This measure was used in order to assess how successfully the methods could satisfy triplets that were not provided. Once the embedding was created from triplet ratings, the Euclidean distances between the points were calculated, range normalized, and converted to a pairwise similarity matrix.

3.4 Measuring similarity directly from products

After the human conceptualization of function was quantified, the results were compared to how quantitative measures determined functional similarity. In order to do this, the full, lower-level functional models (lower image in Figure 1) for the same set of products were represented in a mathematical space as binary matrices, specified using 21 functions and 19 flows (as defined by the functional basis framework). A 1 was used for the existence of a specific function in the product and a 0 was used for the absence of a specific function in the product [14].

The quantitative measures of similarity used were those considered extensively in prior work by the authors [18]: simple matching coefficient, Jaccard similarity, cosine similarity, spectral distance [39], NetSimile [40], and DeltaCon [41]. It should be noted that there are many possible ways to measure similarity and not all were included here. The six measures included in comparison represent six different characterizations of similarity when applied to functional representations of products. The measures range from those that are easily interpretable to those that are not. In addition, although the measures represent more general formulations that have been applied across several domains, efforts were made to select measures that were the most meaningful for the context of engineering design. For example, versions of cosine similarity and a matching measure much like the simple matching coefficient or Jaccard similarity have been applied to engineering design [5–7]. More details on these measures specifically applied to functional models can be found in our previous work [18]. The first three listed measures involve variations of directly matching the existence of features (in this case, functions or flows) across the products being compared. The latter three listed measures involve modeling the products as networks and then comparing the network structure in various ways. For example, the spectral measure incorporates information node degree, which refers to the number of sub-functions operating on a specific flow or a sub-function operating on a number of flows. This could represent the relative importance of specific functions and flows within a functional model. By using two different types of measures, we hoped to include possible connections to *one-to-one matching* and *relational comparison*, both aspects of the structural alignment model of how humans determine similarity [21].

The measures were calculated on the product function matrices using the SciPy, NetworkX [42], and NetComp [39] libraries in Python to obtain pairwise comparisons. These pairwise comparisons were range normalized and then converted to similarities if the original form was a distance or dissimilarity. Therefore, all relative comparisons were scores between 0 and 1, with 1 representing the highest similarity (only found for a product compared to itself) and 0 representing the lowest similarity.

3.5 Comparing human judgements with computed similarity

An additional measure was used for comparing the pairwise similarity scores from the quantitative measures to those obtained from the crowdsourcing study. The comparison of human and computational output was formulated as a search problem: a product was selected as if it was the input of a search and all other products were ranked relative to that product. This was repeated for all products (N=20) and the median was taken to represent results across all products. Typically, rankings can be compared using statistics such as Spearman's rho and Kendall's tau. However, these statistics do not account for the position of the rank differences: a ranking that differs in the highest ranks cannot be distinguished from one that differs in the lowest ranks. As such, it is difficult to separate the effects of similarity and dissimilarity. Therefore, normalized discounted cumulative gain (NDCG) was adapted from the field of information retrieval to compare between the human and computational sources [43]. NDCG can be used to compare rankings to a "ground truth," given relevance scores, with the higher ranks having more importance than lower ranks (rankings can be reversed when considering dissimilarity).

The discounted cumulative gain (DCG) can be found by using a logarithmic discount based on the rank position (i is the rank position, rel_i is the relevance at rank position i, and n is the total number of ranks) as following

$$DCG = \sum_{i=1}^{n} \frac{rel_i}{log_2(i+1)}, \tag{1a}$$

after which it must be normalized by the ideal discounted cumulative gain (IDCG). The IDCG refers to the value of DCG when the list is sorted in order of relevance so that the highest rank has the highest relevance [43].

$$nDCG = \frac{DCG}{IDCG} \tag{1b}$$

In this case, the crowdsourced rankings were considered to be the "ground truth" and each numerical rank from the crowdsourced ranking was used as a relevance score. This was used to calculated the IDCG. Then, the DCG was calculated for the ranking of the measure (e.g. cosine similarity) being compared. Once again, the numerical ranks were used as a relevance score. Similarity thresholds were considered in the analysis to acknowledge that some of the products within the global product space would not be similar to each other at all. When a threshold was set (using the entire pairwise score matrix), products below the threshold then would have a relevance of 0 when appearing in any ranking. If the rankings by humans and a quantitative measure were exactly the same, the NDCG would return a value of 1. NDCG was calculated using the Scikit-learn library in Python.

4 Results

Similarity determined from crowdsourced human data was compared to the calculated similarity scores using the methods outlined in Section 3 and considering levels of abstraction. First, the data collected by having participants consider a lower-level abstraction of product function (i.e. the product's working mechanism) was quantified into a 2D embedding space using t-STE. Other methods for creating the 2D embedding were compared to verify the quality of using the t-STE embedding method. After comparing crowdsourced and computational results at the same level of abstraction, data collected by having participants consider a higher level of abstraction (i.e. the product's purpose) was examined to investigate the effect of abstraction on the agreement between human and computational similarity assessment.

4.1 2D embedding space of product function

The collected triplets regarding a product's working mechanism were used to create a 2D embedding of the product space using t-STE. The final map provides a visualization of which products were considered functionally similar by participants under this perspective and is shown in Figure 3. Before using the pairwise similarity matrix derived from Figure 3 in further analysis, some steps were taken to ensure that the generated embedding provided a satisfactory representation of the human data.

Creation of the 2D embedding was replicated using the three other common triplet embedding methods (GNMDS, CKL, and

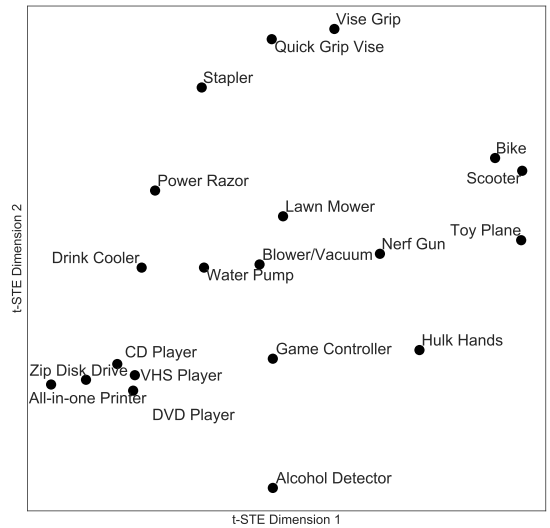
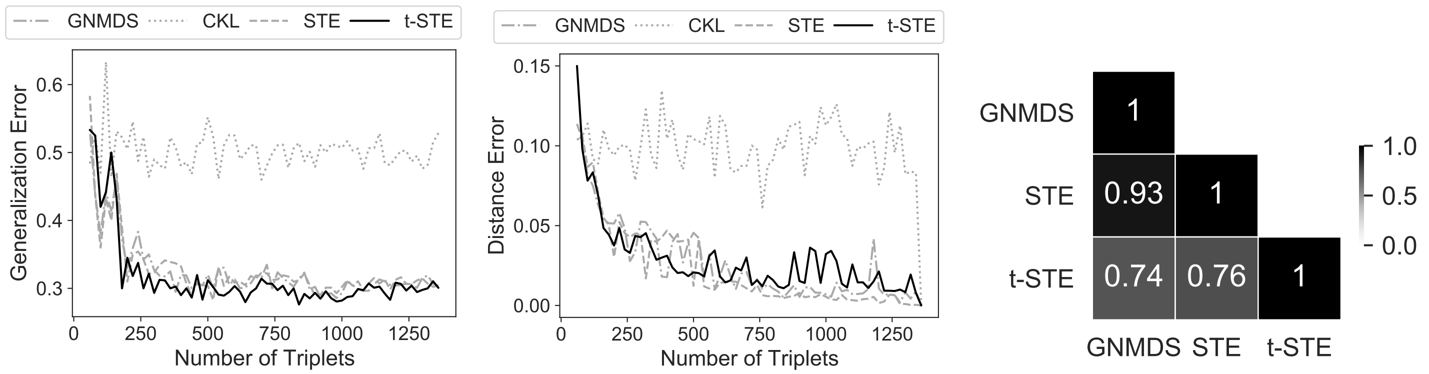


FIGURE 3: Two-dimensional embedding constructed with t-STE from crowdsourced triplets (based on how the products work)

STE). For all of the methods, triplet generalization error and distance error were calculated using fractions of the collected triplets to the full number of collected triplets. As shown in Figure 4a, the GNMDS, STE, and t-STE methods demonstrate a level of convergence before the full number of collected triplets are included. The t-STE method has the lowest triplet generalization error by a small margin when incorporating all of the collected triplets. Even using the full number of collected triplets, about 30 percent of the triplet constraints were not satisfied in the embedding. The occurrence of unsatisfied constraints is in line with previous experiments using triplet embedding methods where not all of the triplet constraints being satisfied [32, 38]. This can be attributed to inconsistency across the crowd among other reasons. Using distance error, as shown in Figure 4b, GNMDS, STE, and t-STE demonstrate a level of convergence in embeddings at about 50 percent of collected triplets. At this point, the similarity scores change only slightly in comparison to the scores from the final embedding. CKL does not demonstrate convergence in either case and therefore, was not considered further.

Finally, the median rank correlation coefficient (Kendall's tau) of the product rankings was calculated across the methods, as shown in Figure 4c, to determine if there were differences in the rankings (i.e. relative similarities calculated from the 2D space) when using a specific triplet embedding method. There is a strong correlation between rankings across methods, demonstrating that in addition to performing similarly in terms of errors in satisfying triplet constraints, the different methods only have a small effect on the resulting pairwise similarities. A closer look at the 2D embedding in Figure 3 verifies that products that were expected to be close to each other in the 2D embedding space (e.g. the two vise grips, located in the upper middle area) are actually close to each other using t-STE.



(a) Average triplet generalization error across ten folds

(b) Mean squared error of distance matrix compared to distance matrix of final embedding

(c) Median rank correlation coefficient between rankings derived from comparable methods

FIGURE 4: Triplet generalization error and distance error verify that the t-STE embedding is not significantly changing with the addition of new triplets after a point. The embedding technique does not significantly impact similarity values derived from the triplet embeddings.

4.2 Human judgements vs. quantitative measures at a lower level of abstraction

Pairwise similarities derived from the embedding shown in Figure 3 (where functional similarity is defined based on how the products works) were compared to those derived from quantitative measures, using the NDCG. When using relative pairwise comparisons, the comparisons must be made using rankings instead of absolute scores since the distribution of values generated across the different similarity measures varies. However, converting to rankings leads to loss of information about whether a product or set of products in the ranking are significantly farther away overall (i.e. the global structure of the 2D space). In other words, since the rankings consider a product compared with all other products, the individual rankings might include products that have very little or no relevance to each other. For this reason, thresholds were introduced to separate relevant products from non-relevant products within the entire product space. The thresholds were based on percentile in order to account for the fact that the same threshold could not be used across all measures. For example, at the 50th percentile, the threshold value was set to the median value of all similarity scores for any given pairwise similarity matrix (crowdsourced or directly computed). Then, when the individual rankings were compared as outlined in Section 3.5, product comparisons with similarity scores that fell below the threshold would be uniformly considered as not relevant for the NDCG calculation.

Similarity. Figure 5 shows the median NDCG across all quantitative measures from using no threshold to a threshold at the 90th percentile (only the top two most similar products are relevant). Though NDCG does not differ by a large amount across any measure, Figure 5 indicates that either the Jaccard similarity or the cosine similarity consistently has the highest median NDCG of all the measures even with varying thresholds. The highest median NDCG corresponds to the highest alignment of a measure with what participants considered to be most functionally similar, at least as a summary across products. Both the Jaccard and cosine similarity measures consider products to be more similar when they have large numbers of features (in this case, function-flow pairs) in common when directly computed on functional models. A high median NDCG for these measures, therefore, may provide some implication that humans also rely on finding and matching common functions or flows when considering how a product works, rather than taking a more holistic view.

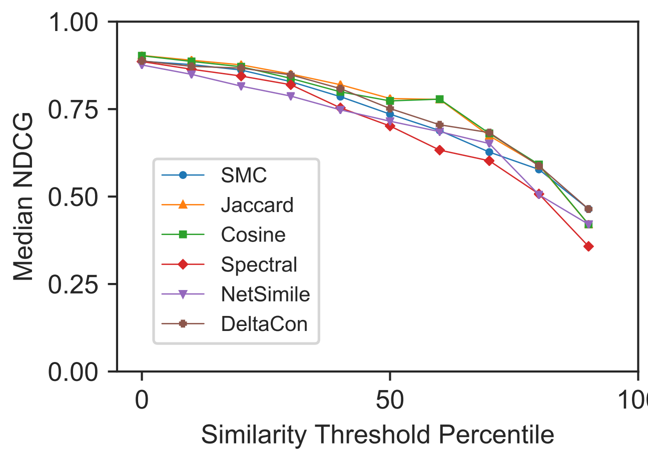


FIGURE 5: Median NDCG for measures compared to human evaluations, considering similar products (based on how the products work)

Dissimilarity. Figure 6 shows results from the same procedure to separate relevant products from irrelevant ones. However, the NDCG now accounts for reversed rankings to prioritize human-computation alignment with regard to dissimilarity rather than similarity. With no threshold, at the 0th percentile, the notion of dissimilarity is based truly on the most dissimilar products. How-

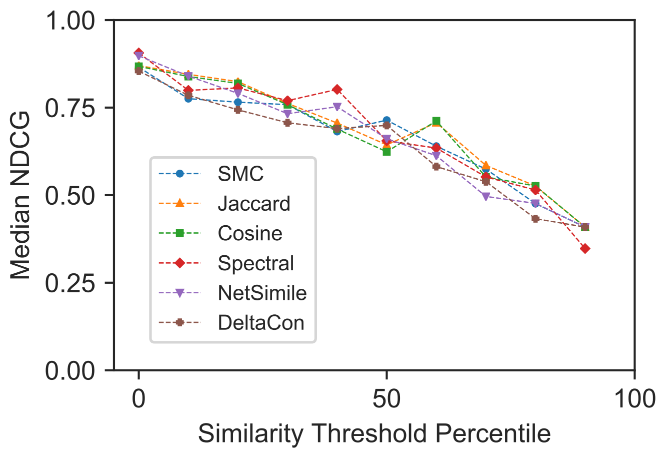


FIGURE 6: Median NDCG for measures compared to human evaluations, considering dissimilar products (based on how the products work). As the threshold increases, dissimilar products may be viewed as "somewhat similar" products.

ever, it should be noted that with higher thresholds applied, the notion of dissimilarity slightly changes: dissimilar products are considered only out of relevant products, as determined by the threshold. Therefore, as threshold values increase, NDCG prioritizes what may be seen as the middle-ground where the products are not too similar, but also not too dissimilar (in design-by-analogy, this may be what has been referred to as the "sweet spot" according to Fu et al. [24]). Figure 6 reveals that there is no measure that has a consistently higher median NDCG when considering dissimilarity. The measure that aligns the most with quantified human judgement highly depends on the line where a product is considered too dissimilar to be relevant. In determining similarity between how products work, therefore, the same quantitative measure may not be able to align with human judgements regarding all three of highly similar products, "somewhat similar" products, and highly dissimilar products.

4.3 Product function at different levels of abstraction

In this section, we probe if the results are affected by changing abstraction level, defined here as the product's purpose vs. working mechanism. Since the participants were presented with the same 20 triplets (in a randomized order) for each level of abstraction, each participant's ratings could be compared across levels to see if the triplets were answered in the same way. Participants answered a median of 70 percent (min: 35%, max: 95%) of the triplets in the same way across both conditions, indicating that participants sometimes had different answers when presented with the two types of function information.

The triplets from when participants were asked to consider the product's purpose (a higher level of abstraction) were then used to create a second 2D embedding of the product space using t-STE. Once again, a pairwise similarity matrix was computed from the embedding. A rank correlation for each product across the levels of abstraction as displayed in Table 1 shows that for some of the products, the rankings match regardless of how function was presented, while for others, the rankings differ significantly. Therefore, it is likely that the non-matching triplets were driven by a smaller subset of products where there is significant divergence in how humans consider similarity when framing function around a product's purpose vs. working mechanism.

TABLE 1: Rank correlation coefficient (Kendall's tau) for each product when comparing human-evaluated similarities across functional abstraction levels. Shaded rows indicate the subset of products with a rank correlation coefficient below the median.

Product	Rank Corr. Coeff.	p-value
Toy Plane	0.31	0.068
Alcohol Detector	0.64	<0.01
All-in-one Printer	0.50	<0.01
Bike	0.65	<0.01
Blower/Vacuum	0.22	0.21
CD Player	0.68	<0.01
Drink Cooler	0.60	<0.01
DVD Player	0.72	<0.01
Nerf Gun	0.40	0.016
Game Controller	0.36	0.034
Power Razor	0.56	<0.01
Stapler	0.58	<0.01
Hulk Hands	0.33	0.049
Lawn Mower	0.42	0.013
Quick Grip Vise	0.73	<0.01
Scooter	0.65	<0.01
VHS Player	0.67	<0.01
Vise Grip	0.78	<0.01
Water Pump	0.18	0.37
Zip Disk Drive	0.59	<0.01

4.4 Human judgements vs. quantitative measures at a higher level of abstraction

Finally, the effect of functional abstraction level was investigated by analyzing the smaller product subset. This subset consisted of products with a low correlation between rankings according to the product purpose vs. how the product works (a correlation coefficient below the median as shown by the shaded rows in Table 1). Then, the same procedure in Section 4.2 was used to create Figure 7.

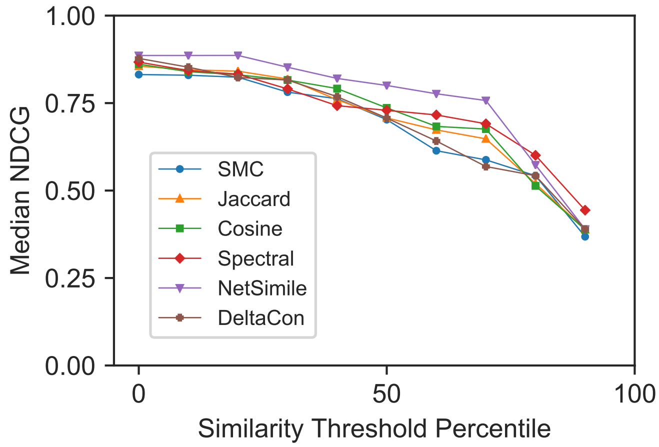


FIGURE 7: Median NDCG for measures compared to human evaluations (based on product purpose) across product subset

Figure 7 shows that for this subset of products, Jaccard similarity and cosine similarity no longer align the most at any threshold. Instead, NetSimile, (a measure that compares feature vectors of network properties) has the highest alignment with humans until a very high threshold, after which the spectral distance (another network-based measure) is the most aligned with the human judgements. Through the lens of dissimilarity, as explained in Section 4.2, no measure clearly has a higher median NDCG across thresholds even at the higher abstraction. Overall, the preliminary evidence from this paper indicates that both the consideration of similarity vs. dissimilarity and higher vs. lower abstraction affects whether a computational measure matches human interpretation of functional similarity.

5 Discussion

In order to apply computational methods to design, it is important to understand when human and computer decisions might be in conflict. In this work, we explore how these conflicts might occur when considering functional similarity by directly comparing results from human judgements to those calculated from functional models. In addition, we identify factors that lead to these conflicts, such as a threshold for similarity or different levels of abstraction. We find that human perception of functional similarity does not necessarily stay consistent when comparing products that are the most similar to each other, as opposed to those that are further away, but still relevant. This is particularly pertinent for analogical design, where comparisons in both the former and latter categories may be important. Additionally, for a subset of products, varying abstraction level considerably affects what people considered to be similar product functions. No single similarity measure matches best with human ratings across abstraction levels, indicating a possible need to use different types of measures (e.g. feature matching or network structure) depending on abstraction. These points are expanded further below.

5.1 Alignment between human evaluations and computed similarity

In the results, we show that crowdsourcing and triplet embedding with t-STE can be used to quantify how people consider products to be functionally similar. We create the human similarity judgement embedding as an aggregate across the participant population though in reality, individuals may perceive similarity in different ways, even when instructions specify consideration along a certain dimension. While many of the limitations to using the 2D embeddings mentioned by Ahmed et al. [32] still apply, the method provides a way to compare these judgements with what we can directly compute from functional representations of products without using a specific design task. In addition, the results support the use of measures such as cosine similarity that have been applied in engineering design, showing them to align relatively well with what humans think when considering high similarity products at a lower abstraction level (how they work). These results are also in line with prior work which indicates that people may use a structural alignment approach in similarity, and specifically notes that people tend to match common features across items, an attribute that is shared by the Jaccard and cosine similarity measure [30].

5.2 Similar, “somewhat similar,” and dissimilar

We find that the measures that align the most in terms of similarity do not necessarily match up in terms of dissimilarity. When looking at the most dissimilar items, there is disagreement between which measures align the most with humans, depending on when an item is included as relevant. Therefore, it appears that the way people consider highly similar items cannot be captured in the same way as how people consider “somewhat similar” items. It is possible that the embedding does not accurately capture how people think of dissimilarity, as they are specifically asked to select the more similar product in the triplet. However, when similarity thresholds are added, these instructions should not affect the results because the dissimilar items are still “relevant” (i.e. similar in some way). There are limitations in the thresholding approach due to the small number of products considered, meaning that certain products may not have had items within the dataset that were similar at all. Further investigation into dissimilarity and similarity thresholds may correspond to finding products that are “far” but not “too far” in terms of analogical distance. Additionally, it is possible that when deciding to utilize a measure to search for far-field sources of inspiration, it is desirable to choose a measure that does not have the highest alignment with human similarity judgements in order to provide unexpected results, as indicated by Fu et al. [12].

5.3 Effect of abstraction level on alignment

Finally, our results indicate that the level of abstraction can confound the return of products that humans consider to be most functionally similar when computing similarity measures directly on product representations. Although we expected that the 2D embeddings would look significantly different for almost all products when considering the different levels of abstraction, it turns out that a smaller subset of products may drive the differences. An overlay of maps of the subset of products with low rank correlations (below the median) is shown in Figure 8.

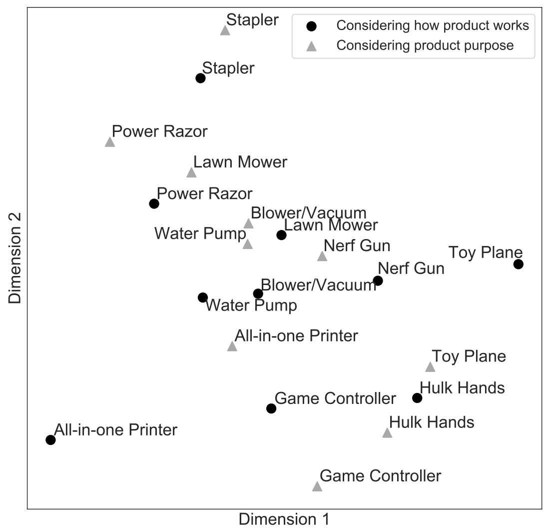


FIGURE 8: Two-dimensional embeddings for the subset of products with low rank correlation coefficients across abstraction levels (visualized using Procrustes analysis in SciPy). Some products are brought closer together or pushed farther apart from each other depending on the level of functional abstraction considered.

From this map, a specific example of the effect of abstraction is in the trio of products including the Hulk Hands, Toy Plane, and Nerf Gun. The Hulk Hands product and the Toy Plane product are closer in the higher abstraction function map (their functions are described as providing sound and motion for entertainment respectively), while the Toy Plane product moves away from the Hulk Hands product and closer to the Nerf Gun product in the lower abstraction function map. This can potentially be explained by the shared pneumatic mechanism between the Toy Plane and Nerf Gun that is not considered for its overall purpose. We note that the function information presented to participants for the lower level of abstraction was summarized from the full functional model and therefore, not necessarily complete. By investigating a larger variety of products, it may be possible to understand the types of products for which abstraction level affects consideration of functional similarity and why. In addition, Chaudhari et al. [44] point out that how people view similarity is dynamic. This is an important consideration when looking at levels of abstraction, where the level of expertise may play a role in the ability to draw more abstract functional connections.

When comparing the human judgements with quantitative measures directly computed on products and including the factor of abstraction, we recognize a discrepancy in access to information: humans were provided with the higher-level function, while the measures still operated on the full, lower-level functional model. This discrepancy can be addressed by using pruning rules on the functional models to remove unimportant information as done by Caldwell and Mocko [26]. However, it may also be desirable for a computed measure to be able to infer the higher abstraction from lower-level attributes rather have to directly provide both levels of abstraction. From this perspective, it is notable that the measures that align the best with human judgements are network-based, in contrast to when considering the lower-level abstraction where more feature-based matching measures suffice. Thus, a network-based measure has the potential to encode the aspect of how humans consider relations and sets of relations within items when making comparisons as proposed by Gentner and Markman [21], in order to allow access to the higher abstraction without the effort needed to directly learn the latent space with large amounts of data.

6 Conclusion

In this paper, we crowdsource human similarity judgements of functional similarity, using a set of consumer products, and apply a triplet embedding method to quantify these human judgments in a 2D embedding space. With this representation, we provide insight into the alignment between how humans view functional similarity and how these functional similarities can be directly computed from the products. We find that measures like cosine similarity, that have been used to calculate functional similarity in prior work, appear to align with the ways in which humans consider highly similar products at lower-level abstraction (i.e. how products work), even outside the context of a particular design task. Our results indicate that the way highly similar products are considered by humans compared to "somewhat similar" products may not be captured by these existing measures, affecting applications such as design-by-analogy, where analogical distance must be controlled. In addition, we find that for some products, the level of abstraction is a factor that can influence whether human judgements align with computational measures. Factoring in higher functional abstraction, network-based measures that account for relations between elements may be appropriate. These types of measures can potentially be used to represent how humans abstract function when it is not possible to directly learn a measure from a large quantity of data collected from humans.

ACKNOWLEDGMENT

This work has been supported by the Regents of the University of California. The findings presented in this work represent the views of the authors and not necessarily those of the sponsors. We would like to thank the authors of [38] for making their code available.

REFERENCES

[1] Christensen, B. T., and Schunn, C. D., 2007. "The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design". *Memory & Cognition,* ***35***(1), Jan., pp. 29–38.

[2] Ulrich, K. T., and Eppinger, S. D., 2004. *Product Design and Development*. McGraw-Hill/Irwin.

[3] Goucher-Lambert, K., Moss, J., and Cagan, J., 2019. "A neuroimaging investigation of design ideation with and without inspirational stimuli—understanding the meaning of near and far stimuli". *Design Studies,* ***60***, Jan., pp. 1–38.

[4] Linsey, J. S., Laux, J., Clauss, E. F., Wood, K. L., and Markman, A. B., 2007. "Effects of Analogous Product Representation on Design-By-Analogy". In DS 42: Proceedings of ICED 2007, the 16th International Conference on Engineering Design, Paris, France, 28.-31.07.2007, pp. 337–338 (exec. Summ.), full paper no. DS42_P_477.

[5] Fu, K., Murphy, J., Yang, M., Otto, K., Jensen, D., and Wood, K., 2015. "Design-by-analogy: Experimental evaluation of a functional analogy search methodology for concept generation improvement". *Research in Engineering Design,* ***26***(1), Jan., pp. 77–95.

[6] McAdams, D. A., and Wood, K. L., 2002. "A Quantitative Similarity Metric for Design-by-Analogy". *Journal of Mechanical Design,* ***124***(2), June, pp. 173–182.

[7] Turner, C., and Linsey, J., 2016. "Analogies from Function, Flow and Performance Metrics". In ICCBR Workshops.

[8] Kittur, A., Yu, L., Hope, T., Chan, J., Lifshitz-Assaf, H., Gilon, K., Ng, F., Kraut, R. E., and Shahaf, D., 2019. "Scaling up analogical innovation with crowds and AI". *Proceedings of the National Academy of Sciences,* ***116***(6), Feb., pp. 1870–1877.

[9] Anandan, S., Teegavarapu, S., and Summers, J. D., 2008. "Issues of Similarity in Engineering Design". In ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, pp. 73–82.

[10] Ranawat, A., and Hölttä-Otto, K., 2009. "Four dimensions of design similarity". In ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE2009, pp. 1069–1077.

[11] Fischer, M. S., Holder, D., and Maier, T., 2020. "Evaluating Similarities in Visual Product Appearance for Brand Affiliation". In Advances in Affective and Pleasurable Design, S. Fukuda, ed., Advances in Intelligent Systems and Computing, Springer International Publishing, pp. 3–12.

[12] Fu, K., Chan, J., Schunn, C., Cagan, J., and Kotovsky, K., 2013. "Expert representation of design repository space: A comparison to and validation of algorithmic output". *Design Studies,* ***34***(6), Nov., pp. 729–762.

[13] Stone, R. B., and Wood, K. L., 2000. "Development of a Functional Basis for Design". *Journal of Mechanical Design,* ***122***(4), Dec., pp. 359–370.

[14] Hirtz, J., Stone, R., Mcadams, D., Szykman, S., and Wood, K., 2002. "A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts". *Research in Engineering Design,* ***13***, Mar., pp. 65–82.

[15] Agyemang, M., Linsey, J., and Turner, C. J., 2017. "Transforming functional models to critical chain models via expert knowledge and automatic parsing rules for design analogy identification". *AI EDAM,* ***31***(4), Nov., pp. 501–511.

[16] Ferrero, V. J., Alqseer, N., Tensa, M., and DuPont, B., 2020. "Using Decision Trees Supported by Data Mining to Improve Function-Based Design". In ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.

[17] Mikes, A., Edmonds, K., Stone, R., and DuPont, B., 2020. "Optimizing an Algorithm for Data Mining a Design Repository to Automate Functional Modeling". In ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.

[18] Nandy, A., Dong, A., and Goucher-Lambert, K., 2020. "A Comparison of Vector and Network-Based Measures for Assessing Design Similarity". In ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.

[19] Vargas, S., and Castells, P., 2011. "Rank and relevance in novelty and diversity metrics for recommender systems". In Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, Association for Computing Machinery, pp. 109–116.

[20] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003. "Metrics for measuring ideation effectiveness". *Design Studies,* ***24***(2), Mar., pp. 111–134.

[21] Gentner, D., and Markman, A. B., 1997. "Structure Mapping in Analogy and Similarity". *American Psychologist*, p. 12.

[22] Goucher-Lambert, K., and Cagan, J., 2019. "Crowdsourcing inspiration: Using crowd generated inspirational stimuli to support designer ideation". *Design Studies,* ***61***, Mar., pp. 1–29.

[23] Srinivasan, V., Song, B., Luo, J., Subburaj, K., Elara, M. R., Blessing, L., and Wood, K., 2018. "Does Analogical Distance Affect Performance of Ideation?". *Journal of Mechanical Design,* ***140***(071101), May.

[24] Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., and Wood, K., 2013. "The Meaning of "Near" and "Far": The

Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output". *Journal of Mechanical Design,* ***135***(2), Feb.
[25] Goel, V., 1995. "Cognitive processes involved in design problem solving". In *Sketches of Thought*. MIT Press, Oct., pp. 95–126.
[26] Caldwell, B. W., and Mocko, G. M., 2011. "Functional Similarity at Varying Levels of Abstraction". In ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, pp. 431–441.
[27] Hebart, M. N., Zheng, C. Y., Pereira, F., and Baker, C. I., 2020. "Revealing the multidimensional mental representations of natural objects underlying human similarity judgements". *Nature Human Behaviour,* ***4***(11), Nov., pp. 1173–1185.
[28] Ellis, D., Whitman, B., Berenzweig, A., and Lawrence, S., 2002. "The Quest for Ground Truth in Musical Artist Similarity.". In Proceedings of the 3rd International Conference on Music Information Retrieval.
[29] Gill, A. S., Tsoka, A. N., and Sen, C., 2019. "Dimensions of Product Similarity in Design by Analogy: An Exploratory Study". In ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.
[30] McTeague, C. P., Duffy, A., Hay, L., Vuletic, T., Campbell, G., Choo, P. L., and Grealy, M., 2018. "Insights into design concept similarity judgements". In 15th International Design Conference, pp. 2087–2098.
[31] Hwang, D., and Wood, K., 2020. "Assessing the Novelty of Design Outcomes: Using a Perceptual Kernel in a Crowdsourced Setting". In Design Computing and Cognition.
[32] Ahmed, F., Ramachandran, S. K., Fuge, M., Hunter, S., and Miller, S., 2018. "Interpreting Idea Maps: Pairwise Comparisons Reveal What Makes Ideas Novel". *Journal of Mechanical Design,* ***141***(021102), Dec.
[33] Siangliulue, P., Arnold, K. C., Gajos, K. Z., and Dow, S. P., 2015. "Toward Collaborative Ideation at Scale: Leveraging Ideas from Others to Generate More Creative and Diverse Ideas". In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, pp. 937–945.
[34] Hope, T., Tamari, R., Kang, H., Hershcovich, D., Chan, J., Kittur, A., and Shahaf, D., 2021. "Scaling Creative Inspiration with Fine-Grained Functional Facets of Product Ideas". *arXiv:2102.09761 [cs]*, Feb.
[35] The Design Repository. http://ftest.mime.oregonstate.edu/repo/browse/.
[36] Tamuz, O., Liu, C., Belongie, S., Shamir, O., and Kalai, A. T., 2011. "Adaptively learning the crowd kernel". In Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11, Omnipress, pp. 673–680.
[37] Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D., and Belongie, S., 2007. "Generalized Non-metric Multidimensional Scaling". In Artificial Intelligence and Statistics, PMLR, pp. 11–18.
[38] van der Maaten, L., and Weinberger, K., 2012. "Stochastic triplet embedding". In 2012 IEEE International Workshop on Machine Learning for Signal Processing, IEEE, pp. 1–6.
[39] Wills, P., and Meyer, F. G., 2020. "Metrics for graph comparison: A practitioner's guide". *PLOS ONE,* ***15***(2), Feb., p. e0228728.
[40] Berlingerio, M., Koutra, D., Eliassi-Rad, T., and Faloutsos, C., 2012. "NetSimile: A Scalable Approach to Size-Independent Network Similarity". *arXiv:1209.2684 [physics, stat]*, Sept.
[41] Koutra, D., Vogelstein, J. T., and Faloutsos, C., 2013. "DELTACON: A Principled Massive-Graph Similarity Function". *arXiv:1304.4657 [physics]*, Apr.
[42] Hagberg, A. A., Schult, D. A., and Swart, P. J., 2008. "Exploring Network Structure, Dynamics, and Function using NetworkX". In Proceedings of the 7th Python in Science Conference, pp. 11–16.
[43] Wang, Y., Wang, L., Li, Y., He, D., Liu, T.-Y., and Chen, W., 2013. "A Theoretical Analysis of NDCG Type Ranking Measures". *arXiv:1304.6480 [cs, stat]*, Apr.
[44] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2019. "Similarity in Engineering Design: A Knowledge-Based Approach". In ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection.

APPENDIX A Products and Function Descriptions

Product	Product purpose (higher abstraction)	How product works (lower abstraction)
Toy Plane	Provide motion for entertainment	Humans pump pressurized air into the plane and throw it to give the plane translational motion. The propellors rotate.
Alcohol Detector	Measure alcohol	Humans turn on the device and blow into it. The device collects the breath sample and uses a chemical reaction to determine and display the alcohol level.
All-in-one Printer	Transform paper	Humans turn the printer on and insert paper. Print data is imported to the printer and then electrical energy is used to signal the printer to release the stored liquid ink. The ink changes the blank paper to the printed paper and the print status is displayed. A scanned document is converted to a signal and exported as scan data.
Bike	Transfer human	Humans pedal to provide mechanical energy for translational motion to transport themselves.
Blower/ Vacuum	Import air and debris and expel air	Humans turn on the device and electrical energy is used to signal the blower or vacuum setting. Air is expelled in the blower setting. An air and debris mixture is taken in and the debris are stored in the vacuum setting.
CD Player	Read a CD	Humans insert a CD and turn on the player. Electrical energy is used to start mechanical rotation of the CD and the lens focuses electromagnetic energy (laser) on the moving disk to read it and play the relevant audio. Buttons are used to control other actions such as pause and repeat.
Drink Cooler	Transfer thermal energy	Humans place the device on a surface and places a cup on top. Electrical energy is used to start mechanical rotation of a fan and extract heat. The fan expels air and the heat is transferred out.
DVD Player	Read a DVD	Humans insert a DVD and turn on the player. Electrical energy is used to start mechanical rotation and the lens focuses electromagnetic energy (laser) on the moving disk to read it. The electromagnetic energy is changed to electrical energy, which is used to display the video and play audio. Buttons are used to control other actions such as pause and eject.
Nerf Gun	Export ammo	Humans load the ammo, pump air into the gun, and pull a mechanical trigger. The pressurized air causes translational motion of the ammo and the gun emits noise.
Game Controller	Control computer	Humans push mechanical buttons or directional joysticks to actuate an electric signal. The electric signal is turned into a control signal that sends data to the connected electronic device as well as into electromagnetic energy (light) and mechanical vibration on the controller.

Power Razor	Separate hair from human	Humans provide translational motion to the razor over the surface of their skin through their hands. Electrical energy is converted to mechanical energy in the razor to cut the hair and separate it from the surface of the skin. The razor releases the cut hair, heat, and noise.
Stapler	Couple paper	Humans store staples in the stapler. Paper is positioned between the top and bottom housing of the stapler and force is applied to the top housing by the hand. The staple is separated from other staples and couples the sheets of paper together. The stapler releases the stapled pages and noise.
Hulk Hands	Emit sound for entertainment	Humans place their hands in the gloves. The gloves detect and process an electrical signal from human movement. The electrical signal is converted to noise.
Lawn Mower	Separate grass from ground	Humans push the lawn mower to add translational motion and turn it on. Liquid fuel is stored and the chemical energy in it is converted to mechanical energy. The mechanical energy is used to cut the grass and expel the cut grass pieces. The lawn mower releases heat, noise, and fumes.
Quick Grip Vise	Secure solid	Humans position the object and secure it by applying force to clamp it.
Scooter	Transfer human	Humans provide or stop translational motion to transport themselves.
VHS Player	Read a VHS tape	Humans turn on the player insert the tape, which is sensed and then guided in. Electrical energy is used mechanically translate the tape and then to start mechanical rotation of the wheels. The magnetic tape reel is read and encoded into video and audio signals, which are played. Electrical energy is also converted to electromagnetic energy (light) to indicate the status. Buttons are used to control other signals such as stop and eject.
Vise Grip	Secure solid	Humans position the object and secure it by applying force, changing its status from unclamped to clamped.
Water Pump	Move liquid	Humans turn the pump on. Electrical energy is converted to mechanical energy and then to pressurized air within the pump, which moves the liquid. Heat, noise, and pressurized air are released.
Zip Disk Drive	Read a zip disk	Humans turn on the reader and insert a zip disk, which is sensed and guided in. Electrical energy is converted to mechanical energy to rotate the disk and to actuate translation for the reading head. The magnetic energy from the disk is converted to electrical energy and is exported as data.