Proceedings of the ASME 2019
International Design Engineering Technical Conferences
and Computers and Information in Engineering Conference
IDETC/CIE2019
August 18-21, 2019, Anaheim, CA, USA

DETC2019-98272

SIMILARITY IN ENGINEERING DESIGN: A KNOWLEDGE-BASED APPROACH

Ashish M. Chaudhari
Graduate Research Assistant
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Ilias Bilionis
Assistant Professor
School of Mechanical Engineering
Purdue University
West Lafayette, IN

Jitesh H. Panchal
Associate Professor
School of Mechanical Engineering
Purdue University
West Lafayette, IN

ABSTRACT

Similarity assessment is a cognitive activity that pervades engineering design practice, research, and education. There has been a significant effort in understanding similarity in cognitive science, and some recent efforts on quantifying the similarity of design problems in the engineering design community. However, there is a lack of approaches for measuring similarity in engineering design that embody the characteristics identified in cognitive science, and accounts for the nature of design activities, particularly in the embodiment design phase where scientific knowledge plays a significant role. To address this gap, we present an approach for measuring the similarity among design problems. The approach consists of (i) modeling knowledge using probabilistic graphical models, (ii) modeling the functional mapping between design characteristics and the performance measures relevant in a particular context, and (iii) modeling the dissimilarity using KL-divergence in the performance space. We illustrate the approach using an example of a parametric shaft design for fatigue, which is typically a part of mechanical engineering design curricula, and test the validity of the approach using an experiment study involving 167 *student subjects. The results indicate that the proposed approach can capture the well-documented characteristics of similarity, including directionality, context dependence, individual-specificity, and its dynamic nature. The approach is general enough that it can be extended further for assessing the similarity of design problems for analogical design, for assessing the similarity of experimental design tasks to real design settings, and for evaluating the similarity between design problems in educational settings.*

Keywords: Similarity assessment, design problems, probabilistic models, causal graphs, KL-divergence.

1 Introduction

Similarity assessment is a fundamental cognitive activity that underlies human judgment, reasoning, learning, and decision making. Similarity is a field of study in diverse domains ranging from cognitive science to visual analytics, and artificial intelligence to Internet search. Since cognitive activities are inherently a part of engineering design, similarity assessment plays an important role in design practice, research, and education.

Designers assess similarity in all phases of the design process. For example, similarity among customers' needs is used to develop product strategies. Designers use functional similarity to generate design concepts. Similarity is the basis for analogy-based design and bio-inspired design. Designers rely on the similarity between models (computational models and physical prototypes) and real designs to predict the actual performance of the product. Similarity between test conditions and the operating conditions in the field is essential for product validation. Similarity with past designs is used for budget allocation in systems design. Algorithms for searching designs (e.g., shape search) are dependent on measures of similarity.

Assessment of similarity is also essential for design research. Researchers use similarity to assess the representativeness of experimental settings relative to real design settings. Similarity aids generalizability of outcomes of experiments to other settings. Similarity of design problems used in human-subject experiments aids in reducing the variability of outcomes of the experiments.

Similarity is also an important consideration in design education. For example, design educators may want to use similar, but non-identical, design problems for students' learning assessment. Sometimes the similarity assessment is between the problems discussed in the classroom and the problems in exams, while at other times, the comparison is with problems in the past exams. Educators are often interested in evaluating the similarity between textbook problems and real-world design problems. Because of its pervasiveness in design-related activities, there is a need to understand *what similarity is, and how to quantify similarity between designs, design settings, and design problems*.

Recently, there have been some efforts on quantifying similarity of design problems. Some of these efforts are geared towards supporting analogy-based design [1, 2], while others are aimed at improving the reliability and replicability of experimental studies [3, 4, 5, 6]. These efforts are focused on identifying the features of design problems that contribute to their similarity. These feature-based approaches are limited in their application to engineering design problems for two main reasons. First, the identified features are appropriate for conceptual design tasks only. There is a lack of approaches to measure similarity of design problems in the embodiment design phase, where scientific (causal) knowledge plays a significant role. Second, existing approaches do not account for the characteristics of similarity, including directionality, context dependence, individual specificity, and the dynamic nature of similarity. These characteristics have been identified by the broader cognitive science research community. To address these limitations, we present a knowledge-based approach for similarity measurement, and show how the proposed approach can be used for engineering design problems.

The paper is organized as follows. We review the literature on similarity assessment in engineering design and in cognitive science in Section 2. The proposed approach for similarity assessment is presented in Section 3. An illustrative example of shaft design for fatigue is presented in Section 4. Section 5 presents the results of an experiment study investigating the relationship between a designer's knowledge level and their usage of the knowledge-based approach versus a feature-based approach. Finally, the potential applications of the proposed approach in design practice, research, and education are presented are discussed in Section 6.

2 Literature Review: Measurement of Similarity

As alluded to earlier, similarity has been a subject of study in many fields. In this section, we review the literature on similarity measurement in engineering design (Section 2.1) because it is the domain of interest for this paper, and foundational research on similarity within cognitive science (Section 2.2) as it has influenced similarity studies in other fields.

2.1 Similarity in Engineering Design Research

Similarity has been studied within engineering design for two main reasons: (i) to support designers in generating new designs, and (ii) to support experimental research in design.

The studies in the former category include design support for analogy-based design. McAdams and Wood [2] present a quantitative metric of similarity for design-by-analogy. The focus is on conceptual design phase and is based on function-based design methodology. Therefore, the similarity metric is based on evaluating functional similarity between products. Fu et al.'s [1] goal is to understand which analogies presented to designers achieve the best design outcomes in analogy-based design. They quantify similarity of designs based on the semantic similarity of patent documents, evaluated using latent semantic analysis. In addition to evaluating the *similarity of designs*, researchers have also relied on *similarity of design problems* to support anaogy-based design. Anandan et al. [7] review similarity metrics from different fields and show how the similarity metrics can be used for evaluating the similarity of designs problems. They argue that similarity of design problems can be defined in terms of the following features: product design specifications, constraints, function structures, concepts, shapes, and manufacturing process plans. McTeague and co-authors [8] attempt to understand how designers make similarity judgments and what aspects of design concepts determine similarity. They view similarity from a structural alignment perspective.

The studies in the latter category focus on finding different design problems that are similar so that they can be used in experimental studies to produce reliable and consistent results. The goal is to achieve improved repeatability through standardization. Durand et al. [5] identify the following structural features of design problems: problem size, functional coupling, participants' familiarity with the design problem/solutions and underlying principles, nature of solution space: size and constraints, effort required to solve the problem, domain of design problem and degree to which analogous solutions can be retrieved. Kumar and Mocko [3] evaluate the similarity of design problems used in research experiments based on their representation. They analyzed 55 design problems from the literature using protocol analysis, and latent semantic analysis. The design problems are characterized based on five structural elements: goals of the problem, functional requirements, non-functional requirements, information about end user, and reference to an existing product. Levy et al. [6] experimentally compare design problems for their similarity in design ideation tasks. The output is measured in terms of student design performance measured using ideation metrics of quantity, quality, novelty, variety, and completeness metrics. Sosa [4] proposes metrics for selecting design tasks in experimental creativity research, which includes design tasks such as idea generation and conceptual design. The phenomena under study are fixation, analogical reasoning, and ideation techniques.

The author [4] examines 160 design tasks from 140 published studies, with different characteristics, including task elaboration, task orientation, task selection, participants, and time allocated for ideation. The metrics proposed are: semantic score, lexical ambiguity, precedent analysis, and readability metrics.

In summary, existing studies on similarity have been primarily focused on the conceptual design phase, with the objective of either supporting designers during the conceptual design or supporting design researchers who study conceptual design. Some of the efforts have been focused on quantifying similarity of designs, whereas others are focused on similarity of design problems. Because of the focus on the conceptual design phase, these studies do not account for the scientific knowledge (e.g., dynamics, thermal science, and mechanics of materials) that designers use to judge the extent to which two designs or design problems are similar (or dissimilar).

2.2 Similarity in Cognitive Science

In cognitive science, the studies on similarity have been focused on descriptive theories of similarity, i.e., understanding how people assess two things to be similar or dissimilar. Typically, these studies start with experiments on human subjects, and then develop models to explain how people judge similarity. The resulting approaches for similarity measurement in cognitive science can be broadly classified into *spatial* approaches and *featural* approaches.

In the spatial approaches, such as in Ref. [9], objects are considered points in a dimensionally-organized metric space, and similarity is inversely related to the distance between objects in the metric space. This approach is suitable for representing features that are naturally described using real numbers.

While the spatial approaches are intuitive, and widely used, they do not account for the **directional nature of similarity**, which has been documented in experimental studies with human subjects. Tversky [10] showed that if there are two objects a and b, then people may assess the similarity of a to b to be different from the similarity of b to a. To address these limitations of the spatial approaches, Tversky [10] developed the featural theory of similarity, where objects are represented as a discrete collection of features, and similarity is directly proportional to the number of common features and inversely proportional to the number of distinctive features. This idea is embodied in the *contrast model*, which uses a linear combination of common and distinctive features. Let A be the set of features of object a, and B be that of object b. Then, the similarity of objects using the contrast model is:

$$S_c(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A), \qquad (1)$$

where $\theta, \alpha, \beta \geq 0$; $A \cap B$ is the set of features common to both a and b; $A - B$ is the set of features that belong to A but not B; and $B - A$ is the set of features that belong to b but not a. Here, f is a non-negative interval scale. Tversky [10] also presented an alternate measure of similarity, called the *ratio model* of similarity:

$$S_r(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)}, \qquad (2)$$

where $\alpha, \beta \geq 0$. Both the contrast model and the ratio model capture the directional nature of similarity.

One of the other characteristics of similarity is that similarity judgments are highly **context dependent** [11]. Two objects may have high similarity in one context, but low similarity in another context. Goodman [12] strongly criticized the notion of similarity. He argued that similarity of two objects is an ill-defined notion because it is not constraining enough. Similarity judgements are not meaningful unless one can say "in what respects" two objects are similar. Different features are important in different contexts. Unless the features and their importance are specified, the similarity cannot be calculated. Medin and coauthors [13] discuss this issue at length. Using the example of plums and lawnmowers, they argued that any two things can be arbitrarily similar and arbitrarily dissimilar [14, pg. 292]. This is because the relative weighting of features and the relative importance of common and distinctive features (the parameters θ, α, β in Equations 1 and 2) vary with the context and the task. In his paper on the featural model, Tversky [10] briefly mentioned that the representation of an object as a collection of features is "a product of a prior process of extraction and compilation". However, he did not provide specifics on that prior process, and how the features can be identified or weighted.

In addition to context dependence, similarity assessment is also **individual specific**. Similarity assessment of individuals depends on their background knowledge, expertise, and past experience. The individual differences in knowledge and expertise result in different representations of problems in the feature space, resulting in individual differences in the perception of similarity and difference [15]. These differences have been shown in the context of physics problems by Chi and coauthors [16].

Finally, knowledge itself is not static; it evolves with experience and learning. Therefore, as the knowledge evolves, the perception of similarity also changes [13]. This **dynamic nature of similarity** has been highlighted by Tversky and Gati [17] – "similarities are constantly updated by experience to reflect our ever-changing picture of the world."

The implication of these characteristics for engineering design is that two designs may be more or less similar, depending on what the designers care about. A human-subject experiment related to a specific design phenomenon may be more or less similar to a "real" design setting depending on the phenomenon that is being studied by the researcher. Similarly, a prototype (or simulation) may be more or less similar to the artifact being designed based on the phase of the design process, and what aspect of the design is being evaluated. The implication of individual specificity is that an expert may see greater similarity between two problems, based on some underlying physics. At the same time, an expert may see greater differences between two problems depending on the deeper differences that may not be apparent to a novice. As a person transitions from being a novice to an expert in a particular aspect of design, his/her perception of similarity also changes.

In summary, the approach for evaluating similarity in engineering design should account for the *causal scientific knowledge* available in the design domain, and the following *characteristics of similarity:* directionality, context dependence, individual specificity, and dynamic nature. In Section 3, we present one such approach where scientific knowledge is used as a basis to account for these characteristics of similarity.

3 A Knowledge-Based Approach for Evaluating the Similarity/Dissimilarity in Design

Our approach for evaluating similarity is based on the hypothesis of Tenenbaum and Griffiths [18] that "a satisfying theory of similarity is more likely to depend upon a theory of generalization rather than vice versa." They argued that "the similarity of y to x may involve the probability of generalizing from x to y or y to x or some combination of those two" [18]. In engineering design, two designs are similar to the extent that one can generalize from one to the other. Indeed, what to generalize depends on what the context is. A test specimen is similar to a real part while attempting to make a design decision, if the outcomes of the test (experiment using the specimen) generalize to the real part. To formalize this intuition let us define some keywords: design, performance, knowledge, and a map between the design and performance.

Design (X). A design X is described by a collection of D parameters which includes factors that may be precisely known and factors that are uncertain but required for evaluating the performance of the design. Given that it is impossible to accurately assess all design parameters the designer places probabilistic beliefs over parameters in the form of a joint probability density, denoted by $f_X(X)$. If the value of a parameter is known, then it basically means that the belief includes a multiplicative Dirac delta factor localized at that value.

Performance (Y). Consider Y to be the P performance parameters that are important in a given context in which similarity is being evaluated. The performance measures embody the context of a design problem because the type of outcomes determines which design parameters are influential and which part of the designer's knowledge is relevant. The similarity between two designs X_1 and X_2 is measured for the given context specified by performance measures.

Knowledge to map design to performance (K). To achieve the desired performance outcomes, a designer requires the knowledge of mapping that is believed to lead to outcomes. This mapping that moves the problem solving from initial design specification to the desired performance is domain-specific and dependent on task environment [19]. It can be described by various forms of knowledge such as causal, topological and procedural knowledge. Since much of this knowledge is causal in nature, especially the physics-based knowledge pertinent to the embodiment design phase, we focus on causal knowledge in this paper. Assume that there are N intermediate variables Z that are required to establish a connection between a design X and its performance Y. We model a designer's knowledge about causal relationships between X, Z, and Y as a directed, acyclic graph K with $M = D+N+P$ nodes. To be consistent, assume that the first D nodes correspond to the design variables X, the next N to the physical variables Z, and the last one to the scalar performance measure Y. We may think of K as a matrix in $\{0,1\}^{M\times M}$, where $K_{ij} = 1$ if there exists a causal link between variable i and j. If no direct causal relationship exists between i and j, then $K_{ij} = 0$. Whether or not a causal relationship exists between i and j depends on whether fixing the former modifies the latter. Since they depend on the designer's prior knowledge and familiarity with the problem domain, the causal relationships remain invariant when external conditions change. It is, however, possible for the lack of prior knowledge to manifest into the epistemic uncertainty over causal relationships.

Mapping design X to performance Y. The causal graph K specifies a factorial decomposition of the joint distribution $f_{Z,Y|X}(z,y|x)$ of the physical Z and the performance measure Y conditioned on the design X. If certain relations between parent and children nodes are deterministic, then this joint distribution uses suitably localized Dirac delta factors. The performance Y of a design X is described by the marginal conditional probability density:

$$f_{Y|X}(y|x) = \int f_{Z,Y|X}(z,y|x)dz. \quad (3)$$

Taking into account that the design X does not fully specify all parameters, we can use Bayes rule to get the the overall belief about the performance parameters:

$$b_{Y|X}(y) := \int f_{Y|X}(y|x)f_X(x)dx. \quad (4)$$

For two designs X_1 and X_2, joint probability densities $f_{X_1}(x)$ and $f_{X_2}(x)$ represent a designer's beliefs about design parameters, whereas $b_{Y|X_1}(y)$ and $b_{Y|X_2}(y)$ represent the designer's beliefs about the performance measure achieved by design X_1 and

X_2, respectively. We propose that a similarity/dissimilarity measure between designs X_1 and X_2 should be defined using a measure of distance between belief states $b_{Y|X_1}(y)$ and $b_{Y|X_2}(y)$.

3.1 Dissimilarity Measure

Based on the above assumptions, we can define the dissimilarity measure of X_1 with respect to X_2 as the Kullback-Leibler (KL) divergence of distribution $b_{Y|X_1}(y)$ from $b_{Y|X_2}(y)$:

$$\begin{aligned} D_Y(X_1,X_2) &:= D_{\mathrm{KL}}(b_{Y|X_1} \parallel b_{Y|X_2}) \\ &= \mathbb{E}_{b_{Y|X_1}}\left[\log \frac{b_{Y|X_1}(y)}{b_{Y|X_2}(y)}\right] \\ &= \int b_{Y|X_1}(y) \log \frac{b_{Y|X_1}(y)}{b_{Y|X_2}(y)} dy. \end{aligned} \tag{5}$$

In the sense of the information theory, the KL-divergence-based dissimilarity measure is interpreted as the amount of (extra) information required to describe the performance of design X_1 using the performance distribution of design X_2, averaged over possible performance of X_1. The higher the information required, the lower is the similarity of design X_1 to design X_2. In the sense of generalizability [18], it represents (inversely) the degree to which the performance of design X_2 generalize to design X_1. The lower the value of $D_Y(X_1,X_2)$, the higher is the generalizability of design X_2's performance to design X_1, and consequently, the higher the similarity of design X_1 to design X_2.

3.2 Properties of the Dissimilarity Measure

To understand why this measure has the prototypical characteristic of dissimilarity, let us discuss some of its mathematical properties:

1. For any two designs X_1 and X_2, we have that the dissimilarity is non-negative, i.e., $D_Y(X_1,X_2) \geq 0$. This makes any two designs comparable.
2. Take the case of two designs that have performance belief states $b_{Y|X_1}$ and $b_{Y|X_2}$ with non-overlapping supports. These designs are clearly dissimilar and indeed the metric gives $D_Y(X_1,X_2) = \infty$.
3. The dissimilarity measure $D_Y(X_1,X_2) = 0$ if and only if $b_{Y|X_1} = b_{Y|X_2}$, i.e., the two designs are indistinguishable if and only if they are believed to produce identical performance outcomes.
4. The measure $D_Y(X_1,X_2)$ is invariant under one-to-one, differential transformations of the performance variable Y. In particular, it does not change if we change the units or scale of Y.
5. Note that the proposed dissimilarity measure is not always symmetric, i.e., $D_Y(X_1,X_2) \neq D_Y(X_2,X_1)$. To understand this property, it is instructive to consider the special case where both $b_{Y|X_1}$ and $b_{Y|X_2}$ are Gaussian, i.e., $b_{Y|X_i} = \mathcal{N}(\mu_i,\Sigma_i)$. Then, we have the following analytical formula:

$$\begin{aligned} D_Y(X_1,X_2) = \tfrac{1}{2}\Big\{ &\mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2-\mu_1)\Sigma_2^{-1}(\mu_2-\mu_1) \\ &-P + \log\frac{|\Sigma_2|}{|\Sigma_1|}\Big\}. \end{aligned} \tag{6}$$

Consider the following subcases:

(a) **Designs with identical performance uncertainty.** Take $\Sigma_1 = \Sigma_2 = \Sigma$. In this case, the metric is:

$$D_Y(X_1,X_2) = \frac{1}{2}(\mu_1-\mu_2)\Sigma^{-1}(\mu_1-\mu_2).$$

Notice that this expression is symmetric $D_Y(X_2,X_1) = D_Y(X_1,X_2)$. In addition, we see that the dissimilarity increases as the distance of the performance means increases. Finally, as the uncertainty of the performance increases (the diagonal elements of Σ become larger) the dissimilarity decreases.

(b) **Designs with identical performance mean, but different uncertainty.** Take $\mu_1 = \mu_2 = \mu$ and, for simplicity $\Sigma_1 = \sigma_1^2 I_P$ and $\Sigma_2 = \sigma_2^2 I_P$, where I_P is the unit matrix in P dimensions. Then the metric becomes:

$$D_Y(X_1,X_2) = \frac{P}{2}\left\{\frac{\sigma_1^2}{\sigma_2^2} - 1 + \log\frac{\sigma_2^2}{\sigma_1^2}\right\}. \tag{7}$$

First, notice that as P increases dissimilarity increases. That is, dissimilarity increases linearly with the number of performance parameters being considered for comparison. Second, the $D_Y(X_1,X_2)$ is clearly non-symmetric. To study this asymmetry, take the function $g(x) = \frac{P}{2}(x - \log x - 1)$, see Figure 1, and notice that:

$$D_Y(X_1,X_2) = g\left(\frac{\sigma_1^2}{\sigma_2^2}\right),$$

and

$$D_Y(X_2,X_1) = g\left(\frac{\sigma_2^2}{\sigma_1^2}\right).$$

It is also easy to show that $h(x) = g(x) - g(x^{-1}) = \frac{P}{2}(x - x^{-1} - 2\log(x))$ has the property that $h(x) > 0$ for $x > 1$. From this property, we get that:

$$D_Y(X_1,X_2) - D_Y(X_2,X_1) = h\left(\frac{\sigma_1^2}{\sigma_2^2}\right) > 0,$$

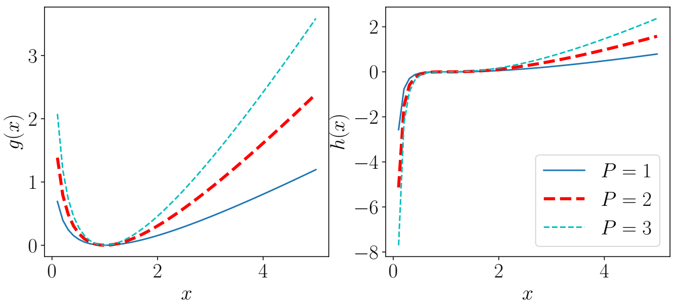


FIGURE 1. A visualization of dissimilarity measure function $g(x)$ and function $h(x) = g(x) - g(x^{-1})$, for designs with identical means but different uncertainty.

FIGURE 2. Similarity is asymmetric; a block on an inclined plane resembles a car on a banked road more than vice versa, both in design parameters and in performance measures. The KL-divergence measure accounts for the asymmetric nature of similarity.

when $\sigma_1 > \sigma_2$. In other words, we have shown that, given the same performance mean, if the performance of design X_1 is more uncertain than that of X_2, then:

$$D_Y(X_1, X_2) > D_Y(X_2, X_1).$$

That is, a high-uncertainty design X_1 is more dissimilar to a low-uncertainty design X_2 than vice versa.
To see why this property is intuitive, consider an example of a block (low-uncertainty design) on an inclined plane versus a car on a banked curve road (high-uncertainty design) in Figure 2. Not only does the car in this situation have more uncertainties in its design parameters such as geometry, weight distribution, contact area etc., than the block, but it also possesses more variance in outcomes such as friction forces. And, it is obvious that we need much more information to describe the results of physical tests on a car using force analysis of a block than to simulate motion of a block given the test results. Therefore, in this context, the car is more dissimilar to the block than the block is to the car. Stated differently, the block is more similar to the car, than the car is similar to the block. This intuition is compatible with Tverskey's [10] feature-based explanation because the car has a greater number of features than the block.

6. Finally, another important property of the KL-divergence is that it is additive for independent distributions. This means that if we have two performance metrics Y_1 and Y_2 that are independent conditional on the design, we get that:

$$D_{Y_1,Y_2}(X_1, X_2) = D_{Y_1}(X_1, X_2) + D_{Y_2}(X_1, X_2). \qquad (8)$$

That is, dissimilarity increases with addition of more independent and distinctive performance metrics. This result is compatible with feature additivity property of feature-based similarity [10] which says that similarity based on two disjoint sets of features is the sum of similarity based on individual sets.

4 Illustrative Example: Shaft Design

In this section, we illustrate the application of the proposed approach for assessing the (dis)similarity using a parametric shaft design problem. We use the approach to identify sets of shaft designs that are similar. The main objective of the example in this paper is to show how the proposed approach satisfies the properties identified in Section 2.2.

4.1 Problem Description

In a shaft design problem, the objective is to design a shaft that can support a set of loads and is safe against yielding and fatigue failure. The shaft is likely to yield due to large load or fail from fatigue due to alternating stresses, and we aim to predict whether such events will occur using the factors of safety.

The geometry of a shaft with two spur gears mounted on it, and supported by two bearings is shown in Figure 3. The geometric parameters include shaft diameters d_1, d_2, and d_3, for sections AD, DE, and EH respectively; gear widths w_1 and w_2; gear pitch diameters, d_{g1} and d_{g2}, of gear 1 and gear 2 respectively; shaft lengths, a, b and L, respectively, between point A and points C, F and H. The boundary conditions B include the simply supported ends that are free to rotate and offer no moment resistance. A point load P_{in} with a constant pressure angle ϕ of 20 degrees acts on gear 1. The rotary motion of the shaft converts the input load to the output load as P_{out} at gear 2. The ultimate tensile strength S_{ut} and yield strength S_y are the material properties. Overall, the design variables in the shaft design problem are $X = \{S_y, S_{yt}, P_{in}, d_1, d_2, d_3, w_1, w_2, d_{g1}, d_{g2}, a, b, L\}$.

We assume that the performance parameters, Y, include either or both of the fatigue factor of safety (FoS) n_f and the yield

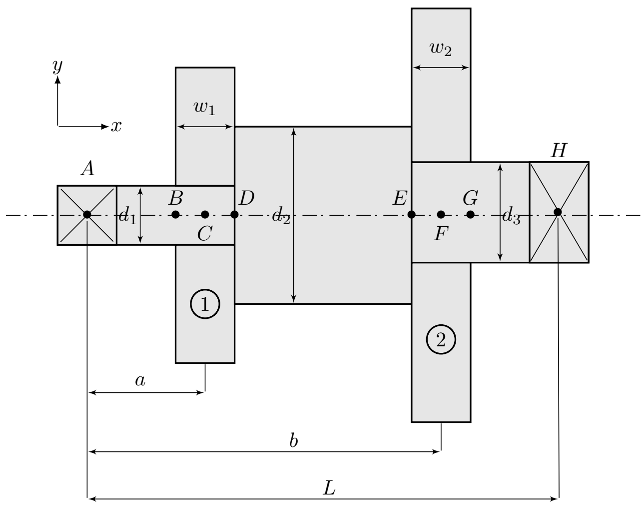


FIGURE 3. Shaft layout for the illustrative example. The shaft has two spur gears mounted on it, and is supported by two bearings at A and H.

FoS n_y.

After defining the design parameters and the performance measures, the details of the knowledge mapping K are in order. Figure 4 summarizes this mapping as a causal knowledge graph. The torque on the shaft is constant $T(t) = T_m = \frac{F_{in}d_1}{2}cos(\phi)$ and leads to nominal shear stress $\tau_0(t)$ in the shaft. The internal load $V(x)$ and internal moment $M(x)$ are non-zero and functions of x, whereas the axial force in the $x-$direction $F(x)$ is zero. Due to continuous rotation, the internal bending moment $M(x)$ results in cyclic bending stress $\sigma_0(x)$ at every point on the outer surface of the shaft.

The nominal shear stress $\tau_0(x)$ due to torsion is a midrange stress and has no alternating component, whereas the nominal bending stress $\sigma_0(x)$ is an alternating stress and has zero mean. Given that $\tau_0(x)$ and $\sigma_0(x)$ have a combined effect on the fatigue life of the shaft, the vonMises stresses for alternating and midrange components at critical sections are written as [20, sec. 6-14]:

$$\sigma'_a = K_f \sigma_0(t) = \frac{32K_f M_a}{\pi d^3}, \tag{9}$$

$$\sigma'_m = \sqrt{3}K_{fs}\tau_0(t) = \sqrt{3}\frac{16K_{fs}T_m}{\pi d^3}, \tag{10}$$

where d is the shaft diameter at the cross-section where the stresses are being calculated. Parameters K_f and K_{fs} are the fatigue stress-concentration factors for bending and shear, respectively. The fatigue stress-concentration factors account for the increased fatigue sensitivity at notches on the shaft. They are included only for transverse cross-sections with immediate vicinity to notches such as at B, D, E and G, and not for cross-sections at C and F.

The theoretical endurance limit S'_e is defined in terms of the ultimate tensile strength S_{ut} for steels based on past empirical observations of fatigue testing [20, sec. 6-7]. The endurance limit S'_e is adjusted through multiplication by *Marin factors* for different conditions of surface finish, size, loading, temperature and miscellaneous factors. The adjusted endurance limit S_e is written as:

$$S_e = (k_a k_b k_c k_d k_e k_f)S'_e. \tag{11}$$

The fatigue strength S_f reduces from S_{ut} for low number of stress cycles N ($N < 10^3$) to S_e for infinite cycles ($N \geq 10^6$). Assuming that the shaft is being designed for infinite number of stress cycles and using the modified Goodman criterion, the fatigue factor of safety n_f and the yield factor of safety n_y for a cross-section of the shaft are written as:

$$n_f = \left(\frac{\sigma'_a}{S_e} + \frac{\sigma'_m}{S_{ut}}\right)^{-1}; \; n_y = \frac{S_y}{\sigma'_a + \sigma'_m} \tag{12}$$

The overall factors of safety for fatigue and yield are minimum of the factors of safety at critical cross-sections such as $B, C, D, E, F,$ and G as shown in Figure 3.

4.2 Simulation Results

To generate shaft design scenarios for similarity assessment, we vary the design parameters X, the knowledge mapping K, and the performance measures Y. Particularly, we assume that a design X consists of deterministic factors that are precisely known and other factors that are uncertain. The known factors are: the yield strength S_y, shaft diameter d_1, and load P_{in}. Accordingly, a point $(S_y \text{ kpsi}, d_1 \text{ in}, P_{in} \text{ lb})$ describes design X. Further, we consider that the designer places belief over the remaining uncertain factors in the form of probability distributions, which are:

$$\begin{aligned} d_{g1}, d_{g2} &\sim \text{Uniform}(2,6); \\ w_1, w_2 &\sim \text{Uniform}(0.5,1); \\ a, (b-a), (L-b) &\sim \text{Uniform}(2,10). \end{aligned} \tag{13}$$

The intermediate variables Z take values according the following relations. The ultimate tensile strength S_{ut} is evaluated from yield strength S_y using an empirical relationship for steels [21]. We have that $d_3 = d_1$ and $d_2 = 1.1d_1$; the Marin factor for surface

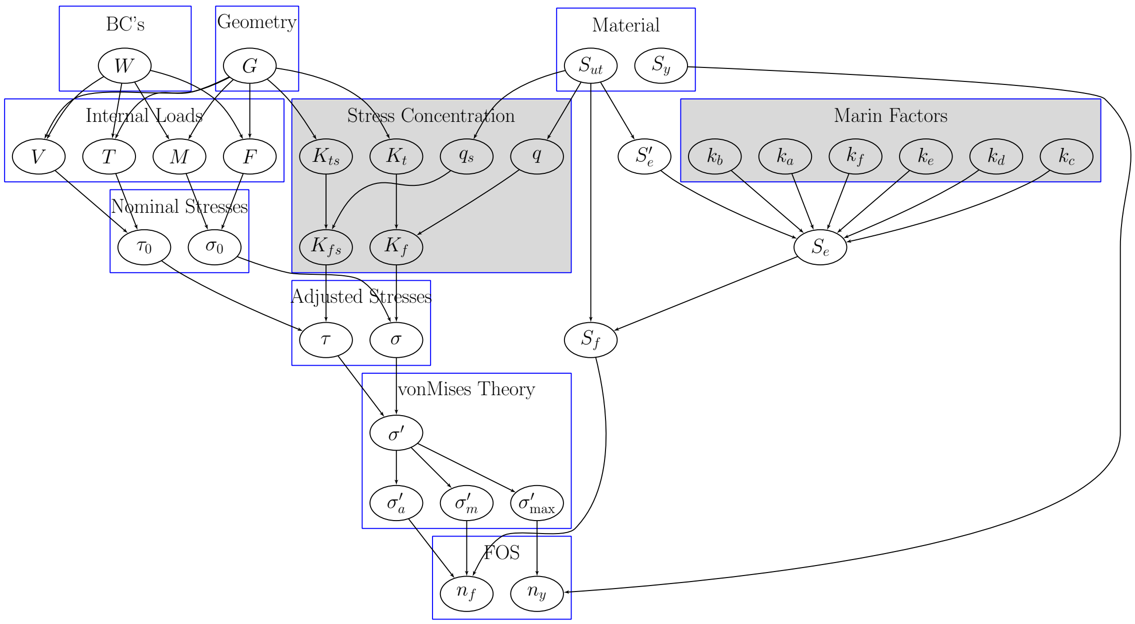


FIGURE 4. A knowledge map for calculating fatigue and yield factors of safety.

finish k_a is 2; the shaft is operated at room temperature, i.e., $k_d = 1$; the reliability factor is $k_e = 0.62$; there are no miscellaneous factors.

Using the above design parameters X and knowledge map K, we perform numerical simulations to approximate the performance distributions $b_{n_f|X}$ and $b_{n_y|X}$. We achieve this by sampling 5000 sets of values for uncertain variables in Eq. (13) and building a histogram of 5000 performance outcomes (n_f and n_y). For example, Figure 5 shows the resulting distributions $b_{n_f|X}$ and $b_{n_y|X}$ for a shaft with $(S_y = 25 \text{ kpsi}, d_1 = 3.5 \text{ in}, P_{in} = 200 \text{ lbs})$, which is labeled design 7, and a shaft design labeled 26 with $(S_y = 125 \text{ kpsi}, d_1 = 3.5 \text{ in}, P_{in} = 950 \text{ lbs})$.

We repeat the above procedure for 27 different designs considering that the yield strength can take values $S_y \in \{25, 75, 125\}$ kpsi, the shaft diameter can take values $d_1 \in \{0.5, 2, 3.5\}$ inches, and the load can assume values $P_{in} \in \{200, 950, 1700\}$ lbs. Table 1 lists the 27 designs with different sets of deterministic factors. For simplicity, we consider that all designs share the same distribution over uncertain factors.

Then, dissimilarity between designs X_i and X_j in the context of performance Y is estimated as the KL-divergence measure between $b_{Y|X_i}$ and $b_{Y|X_j}$. The integration of Eq. (5) is performed numerically. By following this process for 27×27 pairs of designs, we get a dissimilarity matrix $D_Y^{mat} = \{D_Y(X_i, X_j)\}^{27\times 27}$. The dissimilarity between designs is visualized using Multidimensional scaling (MDS) on the average KL-divergence matrix $\frac{1}{2}(D_Y^{mat} + D_Y^{mat\,T})$, where $D_Y^{mat\,T}$ is the transpose of D_Y^{mat}. We make use of `scikit-learn` software [22] for this. The MDS operation places each design on a 2-dimensional space while maintaining the between-design distance specified by the average KL-divergence, as well as possible. The closer the designs are on a multidimensional scaling (MDS) plot, the higher is the similarity between those designs.

Next, we discuss the properties of the dissimilarity matrix in the context of the shaft design example.

4.2.1 Context Effects

Similarity depends on the performance parameters Y we are interested in. The KL-divergence-based dissimilarity measures are different for fatigue factor of safety (FoS) as the outcome $(Y = n_f)$ and yield FoS as the outcome $(Y = n_y)$, see their MDS representations in Figure 7. The relative placement of cluster $(7, 16, 25)$ and cluster $(9, 14, 24)$ on MDS plots changes significantly between n_f and n_y parameters, indicating that the similarity also changes. Designs 7 and 26 are farther apart on the MDS plot of $Y = n_f$ than on the MDS plot of $Y = n_f$. This implies that designs 7 and 26 have greater similarity when the performance is defined by the yield FoS (KL divergence=0.02) as both designs are safe in yield $(n_y > 1)$, as seen in Figure 5. Their similarity reduces when the performance measure considered is the fatigue FoS, because design 26 is more likely to fail under fatigue $(n_f < 1)$ than design 7.

4.2.2 Individual Differences

The measure of similarity depends on the designer's belief about the knowledge map K. Consider that there are two designers Alice and Bob. Alice is an expert who has a high level of knowledge about the domain of shaft design, and adheres to the causal knowledge map from Figure 4 for calculating fatigue FoS n_f. On the other hand, Bob is

TABLE 1. Different values (x_d) of deterministic design parameters, yield strength S_y, shaft diameter d_1, and input load P_{in}, reported in sequence (S_y kpsi, d_1 in, P_{in} lb).

	x_d		x_d
1	(25, 0.5, 200)	15	(75, 2, 1700)
2	(25, 0.5, 950)	16	(75, 3.5, 207.5)
3	(25, 0.5, 1700)	17	(75, 3.5, 950)
4	(25, 2, 200)	18	(75, 3.5, 1700)
5	(25, 2, 950)	19	(125, 0.5, 207.5)
6	(25, 2, 1700)	20	(125, 0.5, 950)
7	(25, 3.5, 200)	21	(125, 0.5, 1700)
8	(25, 3.5, 950)	22	(125, 2, 207.5)
9	(25, 3.5, 1700)	23	(125, 2, 950)
10	(75, 0.5, 200)	24	(125, 2, 1700)
11	(75, 0.5, 950)	25	(125, 3.5, 207.5)
12	(75, 0.5, 1700)	26	(125, 3.5, 950)
13	(75, 2, 200)	27	(125, 3.5, 1700)
14	(75, 2, 950)		

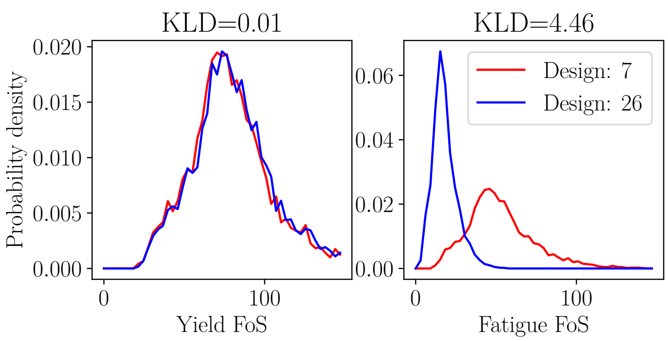


FIGURE 5. Fatigue FoS and yield FoS distributions for designs $7:(25.5\text{ kpsi}, 3.5\text{ in}, 207.5\text{ lbs})$ and $26:(124.4\text{ kpsi}, 3.5\text{ in}, 950\text{ lbs})$

less of an expert, and lacks the knowledge about modifying factors such as stress concentration and Marin factors. The shaded boxes in Figure 4 are missing in Bob's knowledge graph. The results suggest that Alice would conceive the similarity between shaft designs according to the "full knowledge" plot in Figure 6. Bob, without the knowledge of stress concentration and Marin factors, however, would conceive the similarity according to the "partial knowledge" plot in Figure 6.

Two observations are worth noting in this scenario. First, the cluster of designs 7, 16, and 25 appear similar to the cluster of designs 22 and 26 for Bob with "partial knowledge", however, the two clusters are quite dissimilar for Alice with "full knowledge". Second, problem 6 is dissimilar to the cluster of 19, 20, and 21 and nearby problems for Bob, but it is similar to the that of cluster Alice. The reason for these effects is that neglecting the Marin and concentration factors fails to account for the reduced endurance limit and increased stresses due to stress concentration. These multiplicative factors predict the reduction in endurance limit stemming from varying size, loading conditions etc. Without these considerations in Bob's case, the endurance limit is incorrectly considered larger than actual, and the fatigue FoS is therefore larger for different designs of Table 1. For designs 7 and 26 specifically, the uncertainty and the means in fatigue FoS both increase. The increased uncertainty and means reduce the KL-divergence and, therefore, the conceived dissimilarity between designs 7 and 26 decreases.

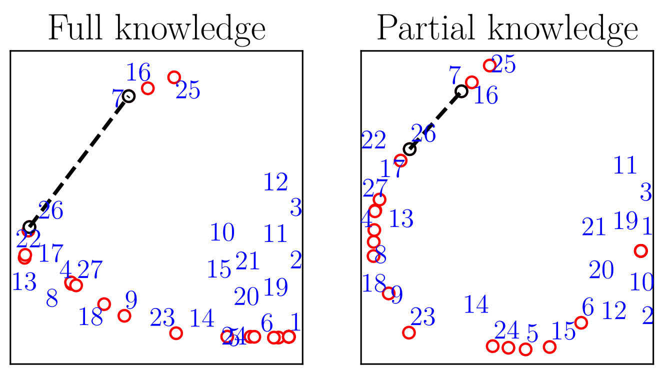


FIGURE 6. MDS representation of the dissimilarity measures with the knowledge of marine factors considered ("full knowledge") and without the knowledge of marine factors ("partial knowledge").

Changing the assumptions about the knowledge map may further change the dissimilarity values. For example, the inclusion of transverse shear stress for FoS calculations, which is ignored in the current causal graph model, will change the fatigue FoS and yield FoS outcomes, especially for a shaft geometry with large diameter and short length. The changed outcomes will result in different MDS representation than the current representations in Figures 7 and 6. Also, the beliefs about the nondeterministic parameters (e.g., the bounds of a uniform distribution in Eq. (13) or the distribution themselves) can be different for different designers.

4.2.3 Directionality

Similarity depends on the direction $X_1 \to X_2$ or $X_2 \to X_1$. We observe that the dissimilarity matrix D_Y^{avg} is asymmetric for all cases of $Y = n_f, Y = n_y$ and $Y = \{n_f, n_y\}$. For a pair of designs which possess asymmetric similarity, this result implies that one design offers better generalizability of per-

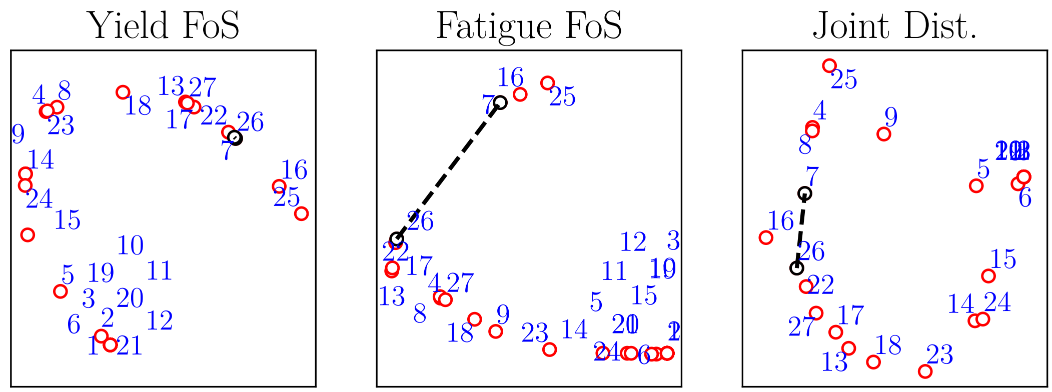


FIGURE 7. MDS representation of the dissimilarity measures for fatigue FoS as the outcome, yield FoS as the outcome, and both factors as the joint outcome.

formance than the other design.

4.2.4 Other Properties

The KL divergence-based dissimilarity measure is also useful for evaluating similarity based on multidimensional outcomes. For a 2-dimensional outcome $Y = \{n_f, n_y\}$, the distribution $b_{Y|X}(y)$ is a joint distribution. We estimate this joint distribution for every design using the sampled values of fatigue FoS and yield FoS. The KL-divergence of joint distributions between two designs is numerically evaluated assuming a 2D mesh grid over the performance space. The multidimensional scaling of these KL-divergences is shown in Figure 7, which is significantly different from those of $Y = n_f$ or $Y = n_y$.

The approach can be used to learn underlying commonalities that are not immediately apparent in the structure space. For example, the proximity of problems $4, 8,$ and 23 on the MDS-based similarity representation in Figure 7 indicates that the effect of increasing the shaft diameter d_1 from 2 inches to 3.5 inches balances the effect of increasing the load P_{in} from 200 lbs to 950 lbs, whereas the same effect is balanced by reducing the yield strength from 125 kpsi to 25 kpsi. As a result, problems $4, 8,$ and 23 are close to each other.

5 Experimental Validation

While the simulation results suggest that the choice of a knowledge graph affects the similarity measure, it remains to be validated whether human designers with different knowledge assess similarity differently. In this regard, the specific simulation results about the individual differences and the context effects lead us to posit the following hypotheses:

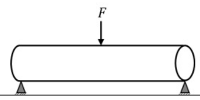
Hypothesis 1: The likelihood of a designer using the knowledge-based approach for similarity assessment increases with their knowledge of the mapping between design and performance.

Hypothesis 2: The likelihood of a designer using the feature-based approach for similarity assessment decreases with their knowledge of the mapping between design and performance.

5.1 Details of the Experiment

To test the hypotheses, we asked the students of an introductory machine design class to answer few questions assessing the similarity between different designs. As shown in Figure 8, each question presented a scenario with the baseline value(s) of design parameter(s), four alternative values of the design parameter(s), and the output parameter upon which the four alternatives are to be compared against the baseline. We designed ten scenarios for similarity assessment in such a way that the prediction of the knowledge-based approach (which compares values of the output parameter) in each scenario would be different from the prediction of the feature-based approach, which compares values of design parameter(s). For example, in Scenario 6 shown in Figure 8, the option with the least Euclidean distance between own design parameters to the baseline design parameters was option (ii). However, option (iv) was closest to the baseline design parameters in the sense of output parameter, i.e., fatigue factor of safety, because both the baseline and option (iv) result in the fatigue factor of safety of 1. This property of the similarity questions provided a direct way to estimate from responses whether a subject followed the knowledge-based approach or the feature-based approach.

The total of ten scenarios were divided into four sets with each set including five scenarios and each scenario repeating in two different sets. Each subject was assigned to a randomly selected set of five scenarios. Overall, 167 subjects responded to the questions asking for the most similar alternatives in part or all of the five scenarios provided to them. The similarity assessment questions were part of the cumulative final exam and included as extra credit questions. The final exam was written and quantitative in nature, and focused on the knowledge of fatigue failure theory[1]. A subject's score in this final exam was assumed as a quantification of their knowledge of fatigue theory concepts.

	Baseline: S_{ut} = 230 kpsi	**Comparison criterion**: S'_e
Scenario 1: A rotating steel shaft with the ultimate tensile strength $S_{ut} = 230$ kpsi is subjected to a constant vertical force F, as shown in the figure below.	(i) $S_{ut} = 70$ kpsi	
	(ii) $S_{ut} = 150$ kpsi	
Baseline: $S_{ut} = 230$ kpsi **Criterion:** Theoretical endurance limit S'_e	(iii) $S_{ut} = 180$ kpsi	**FR**
	(iv) $S_{ut} = 350$ kpsi	**KR**

	Baseline: σ_m = 60 kpsi and σ_a = 40 kpsi	**Comparison criterion**: n_f
Scenario 6: A shaft with tensile strength $S_{ut} = 180$ kpsi and endurance limit $S_e = 60$ kpsi is subjected to fluctuating normal stress with mid-range component $\sigma_m = 60$ kpsi and amplitude component $\sigma_a = 40$ kpsi.	(i) $\sigma_m = 45$ kpsi and $\sigma_a = 30$ kpsi	
	(ii) $\sigma_m = 70$ kpsi and $\sigma_a = 45$ kpsi	**FR**
	(iii) $\sigma_m = 90$ kpsi and $\sigma_a = 35$ kpsi	
Baseline: Fluctuating normal stress with $\sigma_m = 60$ kpsi and $\sigma_a = 40$ kpsi **Criterion:** Fatigue factor of safety n_f for ∞ life	(iv) $\sigma_m = 105$ kpsi and $\sigma_a = 25$ kpsi	**KR**

FIGURE 8. Two example scenarios out of ten scenarios for which the subject population selected the most similar design to the given design. Letters **KR** mark the response as assessed by the knowledge-based approach, whereas letters **FR** mark the response as assessed by the feature-based approach.

5.2 Experiment Results

The results from the experiment suggests a correlation between a subject's knowledge of machine design and fatigue theory and their use of the knowledge-based approach for similarity assessment. Figure 9a shows that the fraction of total responses aligning with the knowledge-based approach increases with the final exam score. Figure 9a also shows 5% and 95% percentile bounds on the fraction of knowledge-based responses at any value of final exam score which are derived from fitting a Gaussian processes regression model on the experimental data. Although the uncertainty bounds seem to depict large variability in the relationship, this relationship is still statistically significant. Fitting a linear regression model between the number of knowledge-based responses and the final exam score provides a slope parameter of 0.007. This slope parameter is statistically significant is at the significance level $\alpha = 0.05$ with t-statistic= 4.63 and p-value< 0.001.

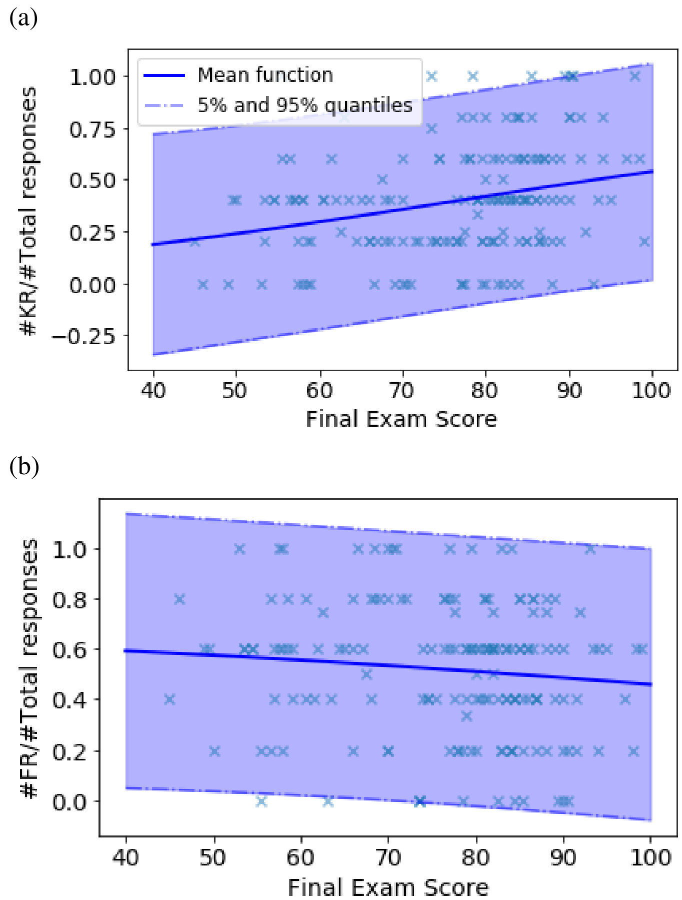


FIGURE 9. a) The fraction of knowledge-based responses versus the final exam score in an introductory machine design class. b) The fraction of feature-based responses versus the final exam score. The data includes responses of 167 subjects. The number of total responses is distributed between one to five as $[5, 1, 2, 10, 149]$.

The results also suggest a trend of decreasing dependence on the feature-based approach for similarity assessment as the level of knowledge increases. According to the observations from Figure 9b, the fraction of total responses aligning with the feature-based approach decreases with the final exam score. This trend, however, is only marginally significant. The slope parameter after fitting a linear regression model between the number of feature-based responses and the final exam score is −0.004. This parameter is marginally significant at significance level $\alpha = 0.05$ with t-statistic= −2.1 and p-value= 0.040.

Further analysis needs to be completed before the hypothe-

[1]The final exam question paper and the questions of similarity assessment are available upon request through email.

ses' validity can be accepted or refuted. Specific directions for future work include building knowledge graphs for individual subjects for systematic and accurate quantification of their knowledge level, quasi-experimental analysis for controlling the effects of external factors, and modeling uncertainty in the subjects' responses.

6 Future Research Directions

The shaft example discussed in Section 4 and experimental validation in Section 5 illustrate the proposed approach, which consists of (i) modeling knowledge using probabilistic graphical models, (ii) modeling the functional mapping between design parameters and the performance metrics relevant in a particular context, and (iii) modeling the dissimilarity using KL-divergence in the performance space. The approach accounts for the causal scientific knowledge available in engineering design, which is of importance in the embodiment design phase. The approach also embodies the characteristics of similarity identified in cognitive science, including, directionality, context dependence, individual specificity, and dynamic nature. Though the shaft design example presented in this paper is simple, the approach can be extended to more general settings discussed in Section 1, such as design research, practice, and education.

Design Research. Consider an experimental research setting where a researcher is interested in comparing an experiment to a real-world phenomenon. The design parameters (X) include treatment conditions related to design expertise, incentives, availability of information sources, the type of design process, etc. They are controlled in experiments but not controllable in real-world situations [23]. The performance measures (Y) under consideration depend on the research study. These measures can be creativity and novelty of ideas for the studies of design fixation [24,25], and the amount of information exchanged or search strategies in the studies of decision making [26]. The knowledge mapping from X to Y may be known if the scientific theory is well developed. In scenarios where the theory is not well developed, the knowledge mapping is itself uncertain, and may vary with the lens used to evaluate the mapping. For this setting, the knowledge-based approach suggest that a fair comparison of an experiment to a real-world situation should explicitly and unambiguously specify the phenomenon being studied and the knowledge mapping being assumed. With this assumption satisfied, quantifying the similarity of a real phenomenon to an experiment becomes synonymous to evaluating the external validity of experimental results.

Design Practice. In a design setting where a designer is interested in comparing a prototype to a real part in practice, design parameters X include various physical characteristics such as geometry, material properties, aesthetics, the level of fidelity [27], etc., which vary between a prototype to a real part. The performance Y is defined by measures such as fatigue life, maximum tensile strength, maximum payload, perceived risk of failure, etc. A well-defined performance measure implies that the problem context is fixed. In some cases, the knowledge mapping may be known, empirically or theoretically. In other cases, the knowledge mapping may not be precisely known due to epistemic uncertainty. For this setting, one of the challenges in evaluating similarity is that the designer has to rely on tests and simulations to evaluate performance. Because tests and simulations can be expensive and time consuming, we need efficient computational methods to calculate the KL-divergence-based dissimilarity measure. The a well-defined parametric shaft design problem presented in Section 3 is an example of this setting. For more complex and ill-structured design problems, we need to account for the impact of design process on the performance. This also points to the larger need of incorporating a model of the design process (in addition to the causal knowledge mapping) into the calculation of similarity.

Design Education. In the context of engineering education, we can use the knowledge-based similarity evaluation for drawing insights on student assessment and developing educational assessment tools. For comparing in-class teaching to tests/exams in disciplinary subjects, the design parameters X mostly relate to technical complexity which is well-specified both in in-class teaching and in exam problems. In-class teaching determines the nature and scope of the knowledge graph, which is tested through exam problems. The performance metrics Y such as grades measure students' problem solving abilities. An exam problem only tests certain parts of the knowledge graph. The knowledge-based approach can help us in evaluating dissimilarity of different exam problems, and identifying problems with high generalizability that can collectively and entirely test a given knowledge graph.

The knowledge-based approach of similarity assessment also highlights the importance of developing knowledge-based and process-based assessment for student design projects. In a setting for comparing different design projects in a project-based course such as capstone design, the design parameters X generally consist of theoretical complexity, problem framing, resources required etc. Commonly used performance metrics Y are outcome-specific and delineated by whether a project gets completed by the deadline, whether a design artifact is fully assembled and functional, etc. However, the knowledge graph that leads to an outcome changes with every design project. The theoretical complexity depends on the specific requirements of an individual design project [28]. In addition to theoretical knowledge, failure to account for a multi-level design process [28], or deviation from a rational design process [29] negatively impacts the design performance. Then, just as similarity is contextual, assessment of students' performance in design projects should be conditional on the specialized theoretical knowledge and a prescribed design process.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the National Science Foundation through NSF CMMI Grants 1662230 and 1841062. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] 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), jan, p. 021007.

[2] McAdams, D. A., and Wood, K. L., 2002. "A quantitative similarity metric for design-by-analogy". *Journal of Mechanical Design,* ***124***(2), p. 173.

[3] Kumar, V., and Mocko, G., 2016. "Similarity of engineering design problems to enable reuse in design research experiments". In Volume 7: 28th International Conference on Design Theory and Methodology, ASME.

[4] Sosa, R., 2018. "Metrics to select design tasks in experimental creativity research". *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, jun, p. 095440621877830.

[5] Durand, F., Helms, M. E., Tsenn, J., McAdams, D. A., and Linsey, J. S., 2015. "In search of effective design problems for design research". In Volume 7: 27th International Conference on Design Theory and Methodology, ASME.

[6] Levy, B., Hilton, E., Tomko, M., and Linsey, J., 2017. "Investigating problem similarity through study of between-subject and within-subject experiments". In Volume 7: 29th International Conference on Design Theory and Methodology, ASME.

[7] Anandan, S., Teegavarapu, S., and Summers, J. D., 2006. "Issues of similarity in engineering design". In Volume 1: 32nd Design Automation Conference, Parts A and B, ASME.

[8] McTeague, C. P., Duffy, A., Hay, L., Vuletic, T., Campbell, G., Choo, P. L., and Grealy, M., 2018. "Insights into design concept similarity judgments". In Proceedings of the DESIGN 2018 15th International Design Conference, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia The Design Society, Glasgow, UK.

[9] Shepard, R., 1987. "Toward a universal law of generalization for psychological science". *Science,* ***237***(4820), sep, pp. 1317–1323.

[10] Tversky, A., 1977. "Features of similarity.". *Psychological Review,* ***84***(4), pp. 327–352.

[11] Goldstone, R. L., Medin, D. L., and Halberstadt, J., 1997. "Similarity in context". *Memory & Cognition,* ***25***(2), Mar, pp. 237–255.

[12] Goodman, N., 1972. "Seven strictures on similarity". In *Problems and Projects*, N. Goodman, ed. Bobs-Merril, New York, pp. 437–447.

[13] Medin, D. L., Goldstone, R. L., and Gentner, D., 1993. "Respects for similarity.". *Psychological Review,* ***100***(2), pp. 254–278.

[14] Murphy, G. L., and Medin, D. L., 1985. "The role of theories in conceptual coherence.". *Psychological review,* ***92***(3), p. 289.

[15] Simmons, S., and Estes, Z., 2008. "Individual differences in the perception of similarity and difference". *Cognition,* ***108***(3), sep, pp. 781–795.

[16] Chi, M. T. H., Feltovich, P. J., and Glaser, R., 1981. "Categorization and representation of physics problems by experts and novices$*''. CognitiveScience$, $\boldsymbol{5}(2), apr, pp.$ 121−−152.

[17] Tversky, A., and Gati, I., 1978. "Studies of similarity". In *Cognition and Categorization*, E. Rosch and B. Lloyd, eds. Lawrence Elbaum Associates, pp. 79–98.

[18] Tenenbaum, J. B., and Griffiths, T. L., 2001. "Generalization, similarity, and bayesian inference". *Behavioral and Brain Sciences,* ***24***(04), aug.

[19] Newell, A., Simon, H. A., et al., 1972. *Human problem solving*, Vol. 104. Prentice-Hall Englewood Cliffs, NJ.

[20] Budynas, R. G., Nisbett, J. K., et al., 2008. *Shigley's mechanical engineering design*, Vol. 8. McGraw-Hill New York.

[21] Webster, S., and Bannister, A., 2000. "Structural integrity assessment procedure for europe–of the sintap programme overview". *Engineering Fracture Mechanics,* ***67***(6), pp. 481–514.

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 2011. "Scikit-learn: Machine learning in Python". *Journal of Machine Learning Research,* ***12***, pp. 2825–2830.

[23] Panchal, J. H., and Szajnfarber, Z., 2017. "Experiments in systems engineering and design research". *Systems Engineering,* ***20***(6), pp. 529–541.

[24] Viswanathan, V. K., and Linsey, J. S., 2012. "Physical models and design thinking: a study of functionality, novelty and variety of ideas". *Journal of Mechanical Design,* ***134***(9), p. 091004.

[25] Toh, C. A., Miller, S. R., and Kremer, G. E. O., 2014. "The impact of team-based product dissection on design novelty". *Journal of Mechanical Design,* ***136***(4), p. 041004.

[26] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2018. "How do designers choose among multiple noisy information sources in engineering design optimization? an

experimental study". In ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. V02AT03A021–V02AT03A021.

[27] Starkey, E. M., Menold, J., and Miller, S. R., 2019. "When are designers willing to take risks? how concept creativity and prototype fidelity influence perceived risk". *Journal of Mechanical Design,* ***141***(3), p. 031104.

[28] Wolmarans, N., 2016. "Inferential reasoning in design: Relations between material product and specialised disciplinary knowledge". *Design Studies,* ***45***, pp. 92–115.

[29] Shergadwala, M., Bilionis, I., and Panchal, J. H., 2018. "Students as sequential decision-makers: Quantifying the impact of problem knowledge and process deviation on the achievement of their design problem objective". In ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. V003T04A011–V003T04A011.

