

Adaptive Inspirational Design Stimuli: Using Design Output to Computationally Search for Stimuli that Impact Concept Generation

Kosa Goucher-Lambert¹

Department of Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
kosa@berkeley.edu

Joshua T. Gyory

Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
jgyory@andrew.cmu.edu

Kenneth Kotovsky

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA, USA
kotovsky@cmu.edu

Jonathan Cagan

Department of Mechanical Engineering
Carnegie Mellon University
Pittsburgh, PA, USA
cagan@cmu.edu

Goucher-Lambert, K., J. Gyory, K. Kotovsky, and J. Cagan, "Adaptive Inspirational Design Stimuli: Using Design Output to Computationally Search for Stimuli that Impact Concept Generation," *ASME Journal of Mechanical Design*, Vol. 142, No. 9, pp. 091401-1-10, 2020. <https://doi.org/10.1115/1.4046077>

¹ Corresponding author.

ABSTRACT

Design activity can be supported using inspirational stimuli (e.g., analogies, patents, etc.), by helping designers overcome impasses or in generating solutions with more positive characteristics during ideation. Design researchers typically generate inspirational stimuli a priori in order to investigate their impact. However, for a chosen stimulus to possess maximal utility, it should automatically reflect the current and ongoing progress of the designer. In this work, designers receive computationally selected inspirational stimuli midway through an ideation session in response to the contents of their current solution. Sourced from a broad database of related example solutions, the semantic similarity between the content of the current design and concepts within the database determine which potential stimulus is received. Designers receive a particular stimulus based on three experimental conditions: a semantically near stimulus, a semantically far stimulus, or no stimulus (control). Results indicate that adaptive inspirational stimuli can be determined using Latent Semantic Analysis (LSA) and that semantic similarity measures are a promising approach for real-time monitoring of the design process. The ability to achieve differentiable near vs. far stimuli was validated using both semantic cosine similarity values and participant self-response ratings. As a further contribution, this work also explores the impact of different types of adaptive inspirational stimuli on design outcomes using a newly introduced “design innovation” measure. The design innovation measure mathematically captures the overall goodness of a design concept by uniquely combining expert ratings across easier to evaluate subdimensions of feasibility, usefulness, and novelty. While results demonstrate that near inspirational stimuli increase the feasibility of design solutions, they also show the significant impact of the overall inspirational stimulus innovativeness on final design outcomes. In fact,

participants are more likely to generate innovative final design solutions when given innovative inspirational stimuli, regardless of their experimental condition.

1. INTRODUCTION

A wide variety of literature demonstrates the impactful nature of inspirational stimuli on design ideation, such as their ability to assist designers in developing solutions with improved characteristics (e.g., increased solution uniqueness and/or feasibility) [1–4]. Additionally, the distance of the inspirational stimulus from the problem domain modulates its impact [5]. Typically, the “distance” of an inspirational stimulus refers to some measure of a stimulus’ proximity to the problem or design space currently occupied by the designer. When viewed on a continuum, the measure of distance of a stimulus is quantifiable using a variety of techniques, such as semantic similarity comparisons or the similarity between functional representations of designs. One can think of a “near” or “close” inspirational stimulus as one that comes from the same or a closely related domain as the problem. Conversely, a “far” stimulus comes from a distant domain. It has also been noted that near stimuli share significant surface level (object) features with the target, while far stimuli share little or no surface features [6]. However, a critical, and currently overlooked consideration impacting these findings is that the relative position of a designer within the design space is not static; it dynamically changes throughout problem solving.

Since the distance of an inspirational stimulus is relative, a design solution evolving during ideation therefore directly determines what constitutes a stimulus as being either near or far. In other words, a far stimulus at the onset of design ideation is not necessarily the same distance as a far stimulus at the midpoint of design ideation. A goal of our research is to develop an individualized tool that enables designers to leverage the full power of inspirational stimuli during

design ideation and problem solving. For this to be the case, such a tool should adapt to the current state of the designer in order to provide a stimulus that reflects the designers' location within the design space. However, most current approaches to selecting design stimuli are not responsive or adaptive to the dynamic state of designers. Typically, stimuli presented during cognitive studies are determined *a priori*.

The work presented in this paper contributes to and advances this ongoing research area by computationally selecting inspirational stimuli based upon a measure of the real-time status of designers' completed activity. First, our approach determines the location of a designer within a larger design space halfway through an ideation session (referred to as their current "state"). To accomplish this, we employ a method of semantic similarity comparisons, computationally comparing the textual work of the designer to a pre-existing database of design concepts collected as part of a prior research study [1]. Using this information, an adaptive intervention is provided to the designer via a stimulus that is either near or far based upon the semantic similarity between all designs within the database and their current design. Thus, the overall goals of this work include: 1) determining whether or not the chosen method of design state detection and adaptive intervention is feasible (i.e., whether a design state can be measured in real-time), quantifiable (i.e., whether textual similarity can be used to provide near and far adaptive stimuli that are significantly different), and perceivable (i.e., can participants distinguish the differences between these categorizations) to designers, and 2) understanding the impact of adaptive stimuli on measurable design outcomes including the novelty, feasibility, and usefulness of design concepts.

1.1 Analogical Reasoning in Engineering Design

Prior work on the role of inspirational stimuli has predominately focused on "analogical reasoning" applied to design. However, there is currently debate as to whether this term is always appropriate

for the design contexts in which it is used [7]. Formally, analogical reasoning is the process of retrieval and mapping of relations or information from a source to a target [8–10]. In this work, the term “inspirational stimulus” is utilized to more broadly encompass other types of stimuli intended to support design ideation that may not satisfy both of these conditions (retrieval and mapping). For example, a prior solution provided to a designer may enhance the likelihood of retrieving a useful concept from memory but does not guarantee a direct mapping incorporating aspects from the stimulus in a new solution for the problem.

The relationship between the distance of inspirational stimuli and solution outcomes is also well studied. One intriguing result is the notion of a “sweet spot” (between near and far) of distance from a stimulus to the problem domain in which a stimulus is most impactful [5]. Because defining a sweet spot for a given research problem is an open area of research itself, most research investigations rely on the comparison between near and far stimuli. Recently, Goucher-Lambert and Cagan analyzed the impact of stimuli distance on the novelty (i.e., uniqueness), feasibility, and usefulness of solution concepts across a wide variety of conceptual design problems from literature [1]. That work revealed that near stimuli improve the usefulness and feasibility of design solutions compared to a control, whereas far stimuli improved the novelty of solutions. Additionally, separate work by Goucher-Lambert et al., utilized functional magnetic resonance imaging (fMRI) to study neural activation patterns underpinning generating design concepts with and without inspirational stimuli of varying distances [11]. In that work, inspirational stimuli defined as close to the problem space activated a unique set of brain regions supporting memory retrieval and solving problems via insight (rather than by analysis) [11]. Across these two studies, closer stimuli were found to more reliably associate with positive ideation outcomes based on both behavioral and neuroimaging data. Other researchers have also found supporting evidence that conceptually

near stimuli may in fact lead to better design outcomes than far stimuli [12]. As an additional contribution, the work presented in the current paper explores the impact of inspirational stimulus distance on design outcomes, building on the aforementioned findings.

1.2 Finding and Applying Design Interventions

When should inspirational stimuli be provided to aid designers? Previous research has demonstrated that interventions are best introduced to problem solvers when there exists an open goal (i.e., when the solver has an understanding of the goal(s) they are trying to accomplish, but have not yet become fixated on a specific solution) [13–16]. Based on this, it would appear that inspirational stimuli should be presented at some point during problem solving before the designer has become fixated or has reached an impasse. However, the difficultly lies in determining the exact moment that someone has reached such an impasse.

Instead of trying to provide a designer with an inspirational stimulus at the correct moment, a different approach allows designers to search for stimuli on their own using structured inputs. One such example are ontology-based frameworks where designers can search for text or image-based stimuli by specifying the object (e.g., chair) and function (e.g., to sit) of their ideas [17–19]. Recently in the Human and Computer Interaction (HCI) community, computational tools have been developed that allow for semi-directed analogy mining. Past approaches at solving this problem included word-embedding models such as GloVe [20], and an analogical search engine by Gilon et al., which looks for distant analogies for specific aspects of a product or design [21]. Additionally, Chan et al. developed an approach termed SOLVENT, which draws on pre-annotations by humans regarding different features of possible stimuli (e.g., purpose, mechanisms, findings) and makes connections based upon semantic representations [22]. While this work is promising, future work in this area is necessary to reduce the burden on designers using these tools

to search for relevant analogous examples. The current approach differs from these past contributions by trying to determine design stimuli based on unstructured rather than structured inputs.

Another approach is to recruit the resources of an expert to help guide a designer towards unexplored areas of the design space. One initial step towards real-time management was an empirical study by Gyory et al., that investigated the characteristics of process management which are most effective for design teams [23]. In that exploratory work, a human process manager oversaw the problem-solving process of a collaborative design team solving a conceptual design problem. The managers tracked the state of the designers within the team, and freely intervened with prescribed stimuli (e.g., design components, select keywords, and/or design strategies) to affect the solving process when deemed necessary. These interventions adapt to a teams' state, since the managers provided stimuli they felt were necessary in reaction to the design teams' activity. Teams that were under the guidance of these process managers significantly outperformed teams that were not in terms of the quality of their solution output. The work by Gyory et al. exemplifies the benefits of real-time management and intervention in design teams. In the current work, we build on this idea further by computationally providing real-time adaptive stimuli through semantic similarity comparisons.

1.3 An Approach to Compare Design Content: Latent Semantic Analysis

In order to conduct the semantic similarity comparisons that determine designers' current state, as well as select the specific adaptive stimulus to provide them, Latent Semantic Analysis (LSA) was used. LSA computes the semantic similarity between text-based corpuses and has been shown to be well suited to a variety of semantic comparisons relevant to design. For example, LSA has been used to quantify the level of semantic convergence in language-based communication between

members in design teams [24,25], uncovering patterns in design repositories such as the US patent database [26], and visualizing the similarity between existing design concepts within a pre-defined design space in a network model [27].

LSA uses singular value decomposition (SVD) for dimension reduction [28]. Within this reduced space, semantic patterns can be uncovered between text-based documents by tracking the co-occurrence of words (represented as vectors). The cosine similarity between document vectors, which analytically computes semantic similarities, varies between zero (if the vectors are completely orthogonal and exhibit no similarity) and one (if the documents are identical). The current work leverages this analytical power of LSA to select design artifacts (inspirational stimuli) semantically near and semantically far from the designer's current concept. The design concept is input as an unstructured description of what the designer believes is currently their best design solution.

1.4 Approaches for Measuring and Evaluating Conceptual Designs

In order to study the impact of the introduced adaptive design intervention, this work relies on evaluations performed by trained expert raters (discussed in Methodology Section 2.4). One such evaluation performed by expert raters is to assess the overall design *quality* of each conceptual design. Design quality is a prominent measure throughout the design literature, with the most common definition of design quality being what Shah, et al., term "a measure of the feasibility of an idea and how close it comes to meet the design specification" [29]. In their popular paper on metrics for ideation effectiveness, Shah, et al., represent quality as both tangible, physical characteristics of a design, as well as the functional, performance metrics, describing the nature of designs. Ahmed et al., provide a similar, but more precise definition for the utility of a design as: "a measure of the designs' performance and can depend on multiple domain dependent factors like

functionality, feasibility, usefulness, impact, investment potential, scalability, etc.” [30]. Some works have used less specific derivatives. For example, Hu and Reid attribute quality to be characteristics of “the physical property, user adoption, and cost-benefit ratio” [31]. Although this list is not exhaustive, it is clear that most, if not all, definitions realize quality as a multi-dimensional construct; some definitions focus on function, some focus on form, and others an amalgamation of the two.

How then, are raters supposed to accurately assess such a metric when attempting to take into consideration, or even deduce, its various sub-dimensions? The subjectivity of measuring quality may very well stem from its dimensional and semantic uncertainty [32]. Furthermore, without a more discrete hierarchy, it is possible for raters to internally weigh the underlying sub-dimensions differently during assessments, leading to yet another source of subjectivity. Motivated by this concern, the current work explores the use of a new measure to represent the overall innovative potential of conceptual designs. Unlike quality, this new measure consists of three distinct sub-dimensions (feasibility, usefulness, and novelty) and directly describes how they should be combined.

While various forms for design quality exist, they undeniably have certain dimensional commonalities. Even from the few definitions mentioned above, both Shah, et al., and Ahmed, et al., consider *feasibility*, or the level at which an idea is physically realizable [33]. Additionally, any design artefact must be able to satisfy its intended goal and meet all the design specifications and engineering constraints. Otherwise, the concept would not be *useful* in any form or function. Evidently, these two sub-dimensions (feasibility and usefulness) are commonly considered factors in the various definitions of design quality, even if not explicitly termed by researchers as such.

Novelty is a less common design metric to associate with quality. Still, many researchers

consider novelty in terms of ideation effectiveness and divergent thinking [29], with a common definition being the uniqueness of a design within a pre-defined set of concepts [34]. Important in ideation applications, novelty can lead to a higher probability of producing higher quality solutions via expanding the design space. In terms of product deployment, the novelty of a design can set products apart from each other, especially when products are similarly effective in their function. Therefore, novelty is a facet of innovation in an increasingly competitive marketplace, and a dimension considered moving forward [35,36].

A previous study correlated quality with each of these dimensions (feasibility, usefulness, and novelty), on a corpus of design concepts originating from a cognitive experiment with 1106 designs [1]. The concepts represented solutions to 4 distinct design problems (electricity: $n = 254$, phone: $n = 290$, joint: $n = 276$, surface: $n = 286$). External evaluators, all Mechanical Engineering graduate students, rated the designs on the metrics of feasibility, usefulness, novelty, and quality, each on a range from zero to two. The interclass correlation coefficient was calculated on a subset of the designs for each design metric separately (*feasibility*: $ICC = 0.77$, *usefulness*: $ICC = 0.65$, *novelty*: $ICC = 0.71$, *quality*: $ICC = 0.50$). All resulted in good or excellent consistency among raters, except for quality, which exhibited only a fair consistency. Correlations between dimensions, e.g. quality/feasibility ($r = 0.43$) and quality/usefulness ($r = 0.73$) were moderate to strong. However, there was no correlation between novelty and quality ($r = 0.04$). The latter result is not surprising, as the authors do not expect novelty, when considered by itself, to represent quality (i.e., novel designs may be poor designs). But as mentioned previously, novelty is still an important dimension for innovation. Thus, a focus of this work is on defining a new, and aptly named, measure for design potential as ***design innovation, I***, which considers the feasibility, usefulness, and novelty of a design concept.

2. METHODOLOGY

To test the feasibility and impact of utilizing LSA to determine inspirational stimuli in response to the current design state of the designer, a human cognitive experiment was developed. The experiment explored the effects of two different LSA-determined distances of inspirational stimuli (near vs. far), as well as a control condition where participants were not provided with any stimulus. Their intermediate and final designs were evaluated across a number of outcome measures of interest, including feasibility, usefulness, and novelty.

2.1 Participants

Sixty-six participants (17 male, 49 female) were recruited for the cognitive study using a call for participation at Carnegie Mellon University and offered \$10 compensation for their participation. All participants read, agreed to, and signed a consent form. Demographics consisted of both university undergraduate and graduate students from a variety of majors and research interests including: Engineering, Fine Arts, Computer Science, and Social Sciences. Data from six participants became corrupted during data collection, and thus excluded from the analysis.

2.2 Experiment Overview

Participants recruited for the cognitive study completed the 30-minute experiment, outlined in Figure 1. For the entirety of the experiment, participants interacted with a graphical-user interface (GUI), coded in MATLAB, which displayed the experiment instructions, the problem statement, and a countdown timer during the problem-solving blocks. After reading through the experiment instructions, participants received the problem statement, which asked them to think of solutions to “*minimize accidents from people walking and texting on a cell phone*” (abbreviated). This

design problem, adopted from work by Miller et al. [26], has been previously utilized by a portion of the current research team in similar concept generation and design ideation tasks [1,11,27].

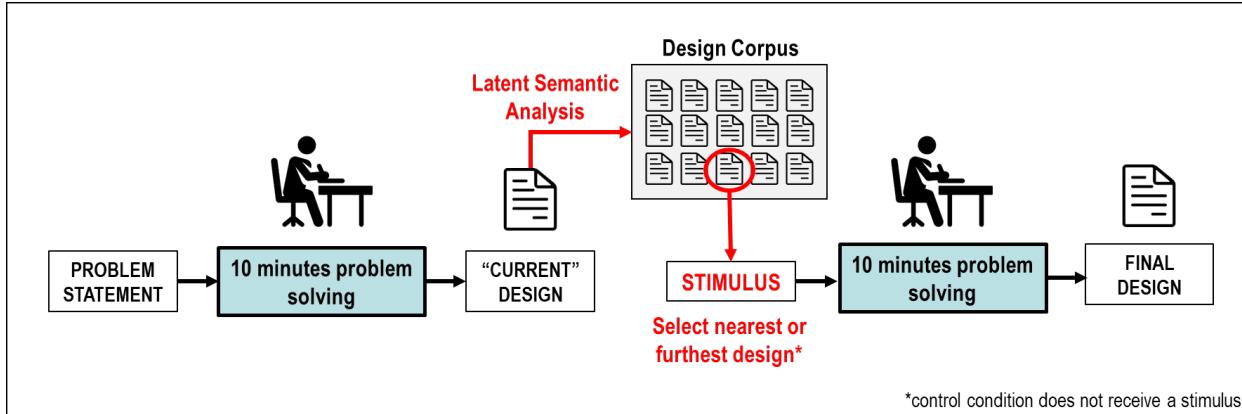


FIGURE 1: OUTLINE OF COGNITIVE STUDY

After reading through the experiment instructions and problem statement at their own pace, participants began generating solution concepts using paper and a digital pen (Neo Lab M1). The digital pen operated in the same way as a traditional pen but tracked pen strokes using a built-in camera (not analyzed in this study). Participants had 10 minutes to ideate and were encouraged to generate as many concepts as they wanted using any combination of textual and/or pictorial representations. At the end of the first 10-minute problem-solving block, the GUI instructed participants to type a 75-word textual description of one design solution in response to the following prompt: “*Please provide what you consider to currently be your best solution*”. Using this textual description of each participant’s current solution, LSA was run to make semantic comparisons between their solution text document and each of the 115 existing stimuli text documents within the design database (described in Section 2.3). The resulting [116 x 116] output matrix from the SVD algorithm was unique to each participant, as it was comprised of both the 115 design stimuli (determined a priori), as well as a participant’s newly developed design.

A balanced experimental design separated participants into one of three experiment conditions: near or far inspirational stimuli, and control, with 20 participants in each condition. Participants only saw one experimental condition during the experiment. The participants in either of the inspirational stimulus conditions (near or far), were immediately, in real time, provided with an inspirational stimulus for review (under 3 seconds of computational time). These stimuli were modulated based on their current design state. For the near inspirational stimulus condition, the stimulus provided was the *closest* stimulus within the design database (115 possible stimuli) to where they were at that point, based on the largest cosine similarity from LSA. In the far inspirational stimulus condition, the *furthest* stimulus within the database was given (lowest cosine similarity). Participants in the control condition immediately transitioned back to ideating after completing the write-up of their best design from the first ideation session. Finally, after the second 10-minute ideation period, all participants completed a write-up of their final “best” design solution. LSA was again performed between this final design write-up and the 115 set of design stimuli for data analysis purposes. The two different LSA comparisons between a participant’s midpoint or final design and the 115 design stimuli allowed for a way to computationally measure the impact of the design stimuli on problem-solving behavior. By computing the semantic distance between participants’ designs and the fixed stimulus space, a sense of the relative movement of a designer within this design space was extracted.

2.3 Design Database

The design database contained 115 possible inspirational stimuli adapted from prior work by Goucher-Lambert and Cagan [1]. During the prior research study, individuals generated 386 solutions for the same design problem employed in the current work. All of the 386 hand-written solutions contained a mixture of text annotations, text descriptions, and drawings. Three

mechanical engineering Ph.D. students, previously trained to evaluate outcome measures (e.g., novelty, feasibility, and usefulness) of the same designs, transcribed descriptions of the content for a random subset of 115 of the 386 design solutions. Each transcription contained a minimum of 75 words. Initial pilot testing identified this word count threshold for each document as being necessary in order to obtain meaningful differences in LSA comparisons. Each of these 115 documents (text descriptions) became one of the potential inspirational stimuli. In the prior work, all 115 inspirational stimuli were evaluated for their novelty, feasibility, and usefulness [1]. Consequently, this study utilized these same rating criteria to make assessments regarding the influence of stimuli on design solution outcomes.

2.4 Analysis of Design Solutions Generated During the Cognitive Study

External raters evaluated both the intermediate designs (D_1 , after the first 10-minute ideation period) and the final designs (D_2 , after the second 10-minute ideation period) on the following outcome measures in order to understand the impact of the computationally selected inspirational stimuli:

1. **Feasibility:** rated on an anchored scale from 0 (the technology does not exist to create the solution) to 2 (the solution can be implemented in the manner suggested).
2. **Novelty:** rated on an anchored scale from 0 (the concept is copied from a common and/or pre-existing solution) to 2 (the solution is new and unique). Of note: “novelty” is considered as the uniqueness of the solution with respect to the entire solution set.
3. **Usefulness:** rated on an anchored scale from 0 (the solution does not address the prompt and/or consider implicit problem constraints) to 2 (the solution is helpful beyond status quo).
4. **Quality:** rated subjectively by each rater on a scale from 0 (low) to 2 (high).

Two trained mechanical engineering Ph.D. candidates, both specializing in design methodology, performed all ratings for solution characteristics. The intraclass correlation coefficient (*ICC*) assessed the consistency between the two design raters using a 25% subsample of the entire dataset. The *ICC* values for novelty (0.78), feasibility (0.65), and usefulness (0.79) were all good or excellent [39]. The *ICC* value for quality was 0.51 (moderate) and therefore excluded from further analysis for being markedly lower than the other measures.

In addition to the metrics noted, participants also provided self-ratings regarding the perceived usefulness and relevance of the provided inspirational stimuli. This information was collected at the end of the experiment, after participants had already written the description of their final design. Participants provided a self-rating for each metric ranging from 1 (low) to 5 (high). The goal in collecting these ratings was to investigate whether or not the computationally determined levels for the inspirational stimuli (near vs. far) aligned with participants' perceptual notion of these categories. Participant self-ratings were not compared to expert evaluations, and therefore a separate scale with a wider range was utilized.

2.5 Design Innovation Measure: A Measure to Assess the Overall Potential of a Design Concept

In addition to assessing the underlying sub-dimensions of the designs (e.g., feasibility, usefulness), it is important to holistically determine an idea's overall potential. To explore the overall "goodness" of the designs in the current research study, the research team adopted the design innovation measure, I , for conceptual design assessment. Design innovation, I , is an encapsulating measure for the overall goodness of a concept, leveraging more well-defined design attributes. These include the feasibility (F), usefulness (U), and novelty (N) of a design, all of which are important for innovation. Accordingly, this measure was defined as follows:

$$I = (F \cdot U) + N . \quad (1)$$

To determine the accuracy and robustness of these underlying sub-dimensions to the overall goodness of concepts, additional formulations of the innovation measure, I , were explored in relation to quality (Table 1). Correlations were run between the design innovation measure and quality, not to necessarily equate the two measures, but to uncover if they follow similar trends.

The variables F , U , and N represent the same sub-dimensions as in Equation 1, and the weights, w_1 , w_2 , and w_3 are determined from a Principal Component Analysis (PCA) run on the rating data. The first formulation placed a greater penalty on feasibility and usefulness as opposed to novelty (i.e., if either F or U scores a 0, the entire ($F \cdot U$) part of that formulation becomes 0). This formulation was motivated by the correlations discussed in Section 1.4 (i.e., feasibility and usefulness being more significant and robust than novelty). Variants two (Equation 2) and three (Equation 3) both assumed equal weighting for the three sub-dimensions, and, consequently, equal importance to innovative potential. However, the second variation allowed for more resolution in the score range and thus also when comparing designs. The multiplicative nature of Equation 3 yielded a larger penalty for scoring zero on any one of the sub-dimensions and introduces a non-linearity for comparable increases to Equation 2. Equation 4 presented a linear combination of the dimensions, weighted by the importance of each in a reduced dimensional space obtained by performing PCA. Justification for the chosen formulation (Equation 1) is presented later in the results section.

TABLE 1: FORMULATIONS OF DESIGN INNOVATION MEASURE, I

$I = (\mathbf{F} \cdot \mathbf{U}) + N$	(1)
$I = \mathbf{F} + \mathbf{U} + N$	(2)
$I = \mathbf{F} \cdot \mathbf{U} \cdot N$	(3)
$I = w_1\mathbf{F} + w_2\mathbf{U} + w_3N$	(4)

3. RESULTS

The resulting data from the methods outlined in Section 2 were analyzed to determine the impact of the computationally adaptive stimuli on design solution output. Specifically, the research objectives include: 1) to determine whether the computational method of design state detection and adaptive interventions via LSA was feasible (i.e., whether the design state can be measured in real-time), quantifiable (i.e., whether textual similarity can be used to provide near and far adaptive stimuli that are significantly different), and perceivable to designers (i.e., can participants distinguish the differences between these categorizations), and 2) to understand the impact of these adaptive stimuli on overall design outcomes (e.g., based on the design innovation measure score), and across final design sub-dimensions, including the novelty, feasibility, and usefulness of solutions.

3.1 Near vs. Far Inspirational Stimuli

The first objective involves determining the utility and validity in using Latent Semantic Analysis to monitor a designer's state. One way to verify the effectiveness of LSA is through examining

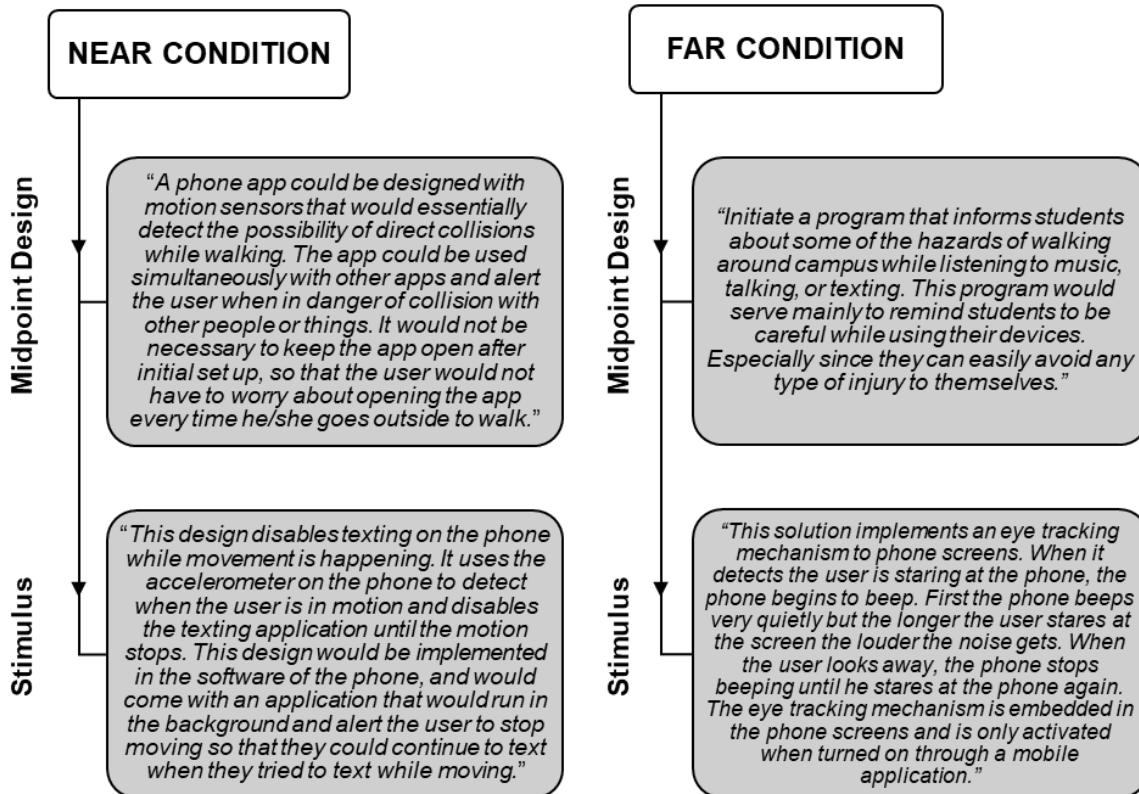


FIGURE 2: EXAMPLE MIDPOINT DESIGNS AND RESPECTIVE PROVIDED STIMULI FOR BOTH THE NEAR AND FAR CONDITION

whether or not the computational approach produced distinct categorizations of the inspirational stimuli provided to the designers. This categorization is determined using the two separate approaches described next. Figure 2 illustrates examples of a participant's midpoint design description and the stimulus they were provided with, in both the near and far condition.

The first approach to verify the effectiveness of LSA compares the average semantic similarity between the stimuli provided in each experimental condition with the state (midpoint design) of that participant. From this analysis, a clear separation between the near and far inspirational stimuli emerges (Figure 3). Near inspirational stimuli have an average cosine similarity of 0.54, whereas

the far stimuli have a much lower similarity of 0.28 ($p < 0.01$). In other words, near inspirational stimuli, as intended, are *much* closer to the (real-time) calculated state of the designer than the far stimuli. This verified that the quantitative method for determining stimulus distance worked appropriately and created substantially difference near vs. far categories.

The second approach leverages ratings-based data collected at the end of the experiment from participants. Each participant was asked how: 1) relevant their provided inspirational stimulus was to their current design on a 1 (not relevant) to 5 (very relevant) Likert scale and 2) helpful their provided stimulus was in developing a solution in response to the problem, again on a scale from 1 (not helpful) to 5 (very helpful). Results indicate that participants perceive near inspirational stimuli as significantly more relevant to their intermediate design solutions than the far stimuli (near: $\mu = 4.225 \pm 0.16$ S.E., far: $\mu = 3.35 \pm 0.36$ S.E., $p < 0.02$, $d = 0.70$). However, participants only found near field stimuli to be marginally more helpful during problem solving compared to the far stimuli (near: $\mu = 3.5 \pm 0.27$ S.E., far: $\mu = 3 \pm 0.35$ S.E., $p < 0.13$, $d = 0.36$). These results validate the computational approach to identify significantly different categorizations of near and

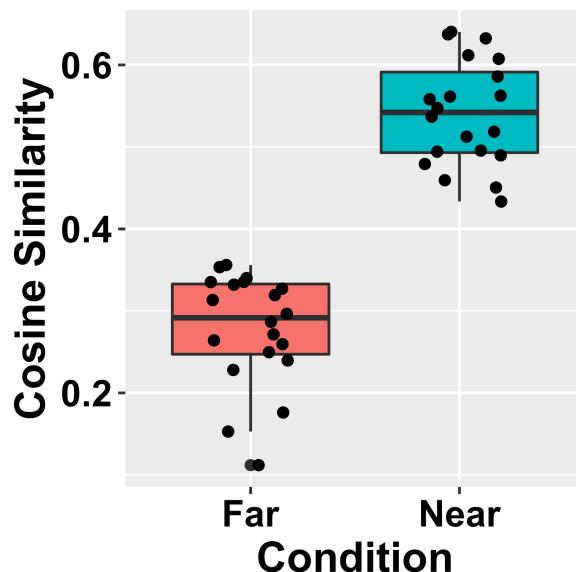


FIGURE 3: THE MEANING OF NEAR AND FAR STIMULI: TWO EXPERIMENTAL CONDITIONS WERE DEFINED USING LATENT SEMANTIC ANALYSIS AND THE CONDITIONS WERE SIGNIFICANTLY DISTINCT FROM ONE ANOTHER (BOXES SHOW UPPER AND LOWER QUARTILES)

far stimuli in response to the current state of designers. Furthermore, these categorizations match the perceived distances of the designers. However, designers perceive both conditions of stimuli as equally helpful to problem solving.

3.2 The Impact of Near vs. Far Inspirational Stimuli on Design Problem Solving

The most important goal of this work involves understanding how these computationally derived stimuli actually affect design ideation. The overall impact of each stimulus on a participant's design output can be measured in a variety of ways. In this work, the two methods employed are: 1) the amount of convergence on the stimulus by the designer and 2) the designer's relative movement within the design space.

The amount of convergence refers to the semantic similarity between the final design and the provided stimulus for both the near and far conditions based upon the LSA cosine similarity value (Figure 4A). From this analysis, results indicate that participants provided with semantically near stimuli converged significantly closer to those stimuli by the end of the experiment ($p << 0.01$, mean cosine similarity values: near: $\mu = 0.33$, far: $\mu = 0.13$). However, while participants remain more similar to near stimuli at the end of the experiment, far stimuli may have had a larger impact in the amount of participants' "movement" within the design space. A relative measure of the overall distance was determined by calculating both the semantic similarities between participants' first design and the stimulus, as well as the final design and the stimulus, and taking the difference between them (Figure 4B). The distances were calculated relative to the design stimuli themselves, because in order to compare this distance across participants, there needed to be a common reference point across participants. The design stimuli served as these common points of reference within the design space. From this analysis, there is a significant difference between the two

conditions ($p < 0.016$). Participants provided with far inspirational stimuli move a greater distance in the design space from the beginning to the end of the design ideation period (near: $\mu = 0.07$, far: $\mu = 0.12$).

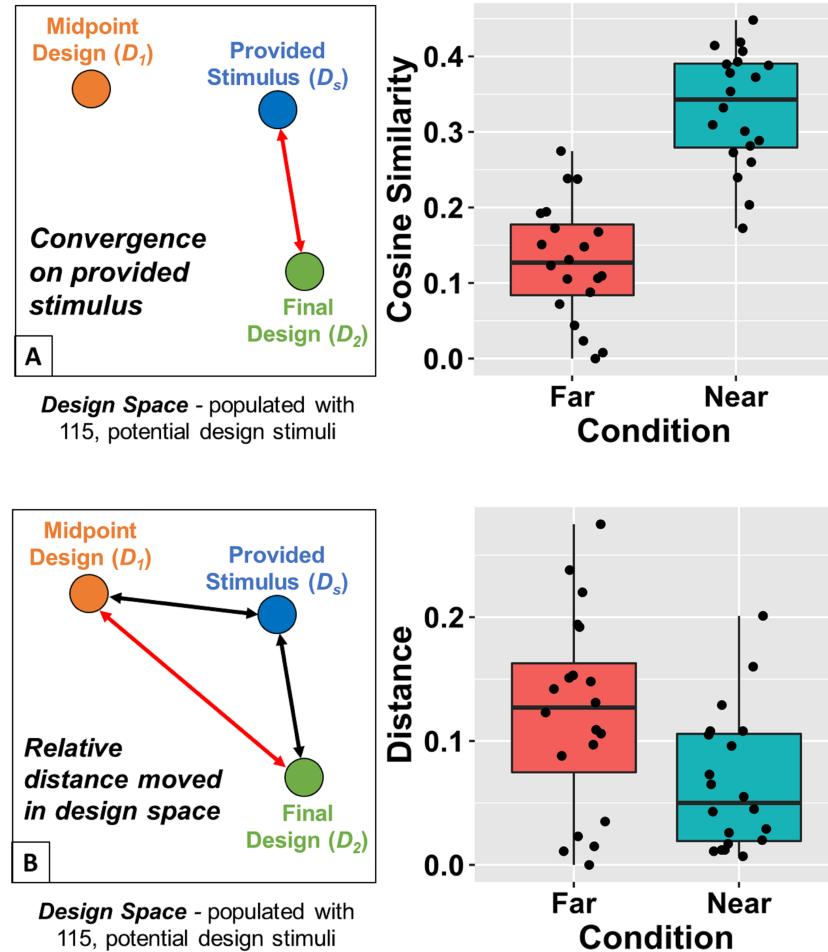


FIGURE 4: THE IMPACT OF INSPIRATIONAL STIMULI ON FINAL DESIGNS. THE LEFT SIDE OF FIGURES 4A AND 4B CONCEPTUALLY ILLUSTRATES THE DISTANCE MEASURED IN THE ACCOMPANYING PLOTS. A) PARTICIPANTS' FINAL DESIGN CONCEPTS ARE MORE SIMILAR TO THE STIMULUS WHEN PROVIDED A NEAR STIMULUS; B) FAR INSPIRATIONAL STIMULI LED TO MORE RELATIVE MOVEMENT (I.E., DISTANCE) WITHIN THE DESIGN SPACE.

3.3 The Impact of Near vs. Far Inspirational Stimuli on Sub-Dimensions of Final Designs

In order to understand the impact of different types of computationally derived inspirational stimuli on sub-dimensions of final designs (novelty, feasibility, and usefulness), the expert ratings of these

design metrics are used. As discussed previously, two trained experts evaluated both the midpoint (D_1) and final design solution (D_2) concepts produced by each participant during the cognitive study. To completely understand the impact of the stimuli on performance, one needs to consider where a designer ended up (i.e., their D_2) in reference to where they started prior to an intervention (i.e., their D_1). By analysing performance in this manner, one can see if providing a near or far stimulus is beneficial or detrimental to problem solving. Using these ratings, the difference between the final design and the prior design is calculated separately for each of the sub-dimension metrics (novelty, feasibility, usefulness), and each experimental condition (Figure 5). Results indicate that there is no significant difference between the conditions in terms of novelty. In other words, providing a participant with a near, far, or no stimulus did not significantly increase or decrease the rarity of their designs from D_1 to D_2 . However, intervening with semantically near inspirational stimuli significantly increases the feasibility of designs compared to providing no stimulus ($p = 0.05, d = 0.5$). Additionally, providing semantically far stimuli significantly decreases the usefulness of designs ($p < 0.01, d = 1.1$).

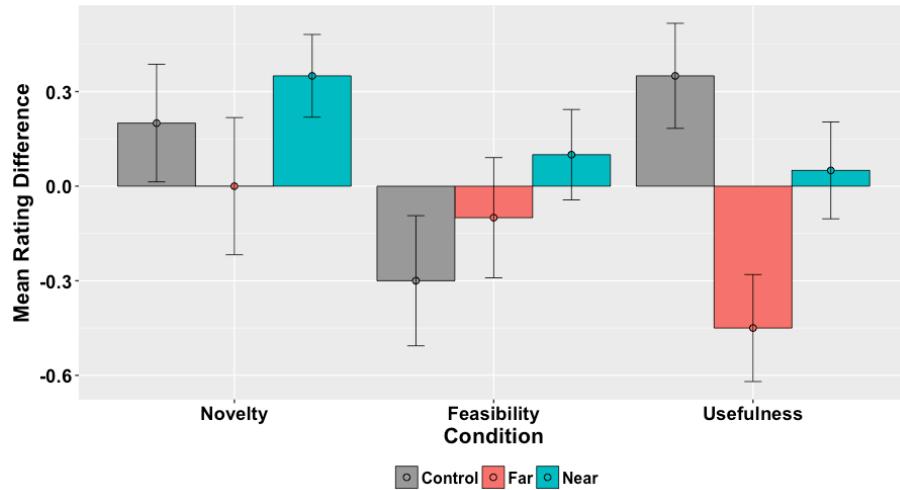


FIGURE 5: MEAN DIFFERENCE BETWEEN FINAL (D_2) AND INTERMEDIATE (D_1) DESIGNS FOR THREE METRICS, SEPARATED BY EXPERIMENTAL CONDITION

3.4 Exploring Multiple Formulations of the Design Innovation Measure

Correlations between design quality and design innovation for the different formulations of the design innovation measure (Table 1) are presented in Table 2. The results in Table 2 are separated based upon the 4 distinct design problems explored in prior work [1]. Specifically, these include: “*A device to remove the shell from a peanut in areas with no electricity*” (electricity), “*A way to minimize accidents from people walking and texting on a cell phone*” (phone), “*A device to immobilize a human joint*” (joint), and “*A device that disperses a light coating of a powdered substance over a surface*” (surface) [3,37,40,41]. The average sub-dimensional weights found through Principal Component Analysis (which was run separately on the three different design problems) was: $w_1 = 0.77$, $w_2 = 0.54$, and $w_3 = 0.37$. Overall, the chosen formulation of Equation 1 shows the highest correlation between innovation and quality on all 4 design problems. Along with its simplicity, Equation 1 is the chosen design innovation measure:

$$I = (F \cdot U) + N . \quad (1)$$

However, it should be noted that all formulations exhibit fairly strong correlations, which supports both the accuracy and robustness of the three chosen sub-dimensions. Furthermore, when taken together, they are representative of the overall goodness of designs and possess merit as a useful design measure, without the burden of the less consistent quality evaluation.

TABLE 2: CORRELATIONS BETWEEN QUALITY AND FOUR FORMULATIONS OF THE DESIGN INNOVATION MEASURE

Equation Problem	(1)	(2)	(3)	(4)
Electricity	0.60	0.56	0.46	0.57
Phone	0.80	0.74	0.71	0.73
Joint	0.72	0.58	0.044	0.71
Surface	0.64	0.60	0.40	0.62

3.5 The Impact of Inspirational Stimuli on Overall Final Design Innovation

An alternate method to examine the impact of the inspirational stimuli is to examine whether or not a specific intervention led to an *overall* better final design (instead of focusing on sub-dimensions as described previously). Recall that one of the design metrics originally rated by the external evaluators was the quality of designs. However, the *ICC* value for quality was much lower than the other design metrics (*feasibility*: $ICC = 0.65$, *novelty*: $ICC = 0.78$ *usefulness*: $ICC = 0.79$, *quality*: $ICC = 0.51$). Consequently, due to this inconsistency between raters, quality cannot serve as a consistent measure of impact of the inspirational stimuli. Instead, the design innovation measure I is used to holistically encapsulate the goodness of design concepts.

Utilizing the design innovation measure, the overall innovative potential of the design stimuli, $I(D_S)$, and both participants' intermediate and final designs ($I(D_1)$ and $I(D_2)$, respectively) are calculated (Equation 1). Similar to the analysis presented previously (Figure 5), performance is

assessed by examining the difference in innovation scores between D_1 and D_2 . Again, the results support a similar finding, with no significant effect in the change of innovation scores (either increase or decrease) in relation to the stimulus condition.

However, it is not completely accurate to only consider whether a stimulus is near or far when measuring its impact on problem solving. One limitation to the above analysis is that it does not consider the quality of the provided stimuli. For example, if a participant received a poor quality stimulus, regardless of its distance, one should not necessarily expect their final design to improve. Because quality is hard to consistently assess, the design innovation measure is again used to represent the overall goodness of stimuli. Here, the different stimuli that designers received during the experiment varied significantly in regard to this measure. Consequently, a follow-up analysis is performed, which considers the innovation score of the stimuli:

$$\text{Corr} ([I(D_2) - I(D_1)], [I(D_s) - I(D_1)]). \quad (5)$$

Equation 5 measures the correlation between a participant's final design innovation score ($I(D_2)$) and the innovation score of the received stimulus ($I(D_s)$), both in reference to their intermediate design solution ($I(D_1)$). From this analysis (Figure 6), it can be seen that an inspirational stimulus with a higher innovation score, relative to the intermediate design, is significantly correlated with a better final design (i.e., an increase in I from D_1 to D_2 ; $r(38) = 0.67, p < 0.001$). To ensure this correlation was independent of the bias introduced via the transformation in Equation 5, an additional analysis was performed. First, 1000 random samples for each of $I(D_1)$, $I(D_2)$, and $I(D_s)$ were drawn from a uniform distribution and fed through Equation 5 (r_{bias}). Next, 1000-tuple samples (each tuple set containing an $I(D_1)$, $I(D_2)$, and $I(D_s)$) were taken from the experimental data (with replacement) and fed through Equation 5 (r_{exp}). A Fisher z-transformation showed that the empirically derived correlation value (r_{exp}) was stronger and significantly different from r_{bias}

($p << 0.001$), revealing the independence of the result from any introduced bias. Using the aforementioned distribution to determine r_{exp} , the following correlation value and 95% confidence interval were obtained: $r_{exp} = 0.67$, 95% CI [0.63, 0.70].

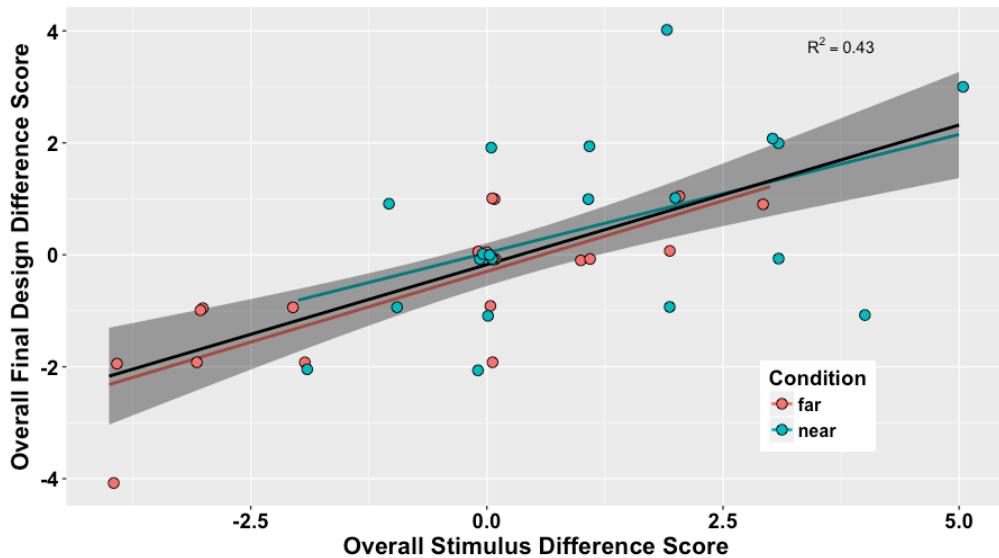


FIGURE 6: THE OVERALL INNOVATIVENESS OF THE PROVIDED INSPIRATIONAL STIMULI IS PREDICTIVE OF THE FINAL DESIGN OUTCOME

The positive correlation between the innovation score of the stimulus and final design is true regardless of whether a participant received a near or far stimulus. Conversely, a participant that received a less innovative stimulus is more likely to produce a less innovative final design. These results provide an interesting, tangential perspective. When intervening during problem solving via inspirational stimuli, adapting stimuli based on relative quality (as measured here by design innovation, I) highly correlates with final design outcomes.

5. DISCUSSION

Overall, this work provides an initial step toward real-time intervention during engineering design problem solving. Unlike other work that provides inspirational stimuli for designers *a priori*, the methods introduced here respond dynamically to an evolving state of participants' design output.

These stimuli are related design solutions sourced from a pre-existing database; they are intended to increase designers' ability to recall useful ideas from memory that may aid in their ability to generate solutions with increased positive characteristics (e.g., feasibility and novelty; see discussion in Section 1). In this work, interventions are provided midway through problem solving and adapted to designers' current solution output. The adapted stimuli, determined using Latent Semantic Analysis (LSA), represented solutions either semantically near or semantically far from designers' present solutions. Thus, these stimuli occupy positions in the semantic design space either close to or far from the designers' relative location.

The results from this study demonstrate the applicability of semantic similarity measures, such as LSA, for identifying stimuli based on the current state of a designer. When the semantically near and far stimuli are extracted from the design space, two distinct and significantly different (in terms of semantic similarity) clusters emerged. This supports the notion that computationally defined near (or far) stimuli are, indeed, near and far. Furthermore, it provides evidence that the design space of stimuli contained designs distinct enough from one another. If the design space did not contain sufficiently distinct designs, it would not have been appropriate to categorize the designs as near and far. Additionally, results from the qualitative analyses showed that participants in each condition perceived the stimulus provided to them as equally helpful. Participants self-rated near stimuli as significantly more relevant to the design problem compared to far stimuli, but not significantly more helpful. Therefore, participants perceived the inspirational stimuli as equally helpful, but not significantly different in terms of their relevance.

While this work applied LSA to adaptively select inspirational stimuli, there are other possible approaches. LSA bases its comparisons on semantic similarity; as such, this method can only handle and compare *textual* outputs of designers. Other topic modelling methods to handle text-

based comparisons are Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA) [42,43]. The work presented in this paper demonstrates LSA's ability to quickly find relevant stimuli for use by designers. However, LSA should be compared to other approaches for adaptively finding textual stimuli for designers. Furthermore, theoretically, there is nothing preventing similar vector-based similarity comparisons from being made between images or other modalities of stimuli. Future work should also consider additional modalities of inspirational stimuli and ways to logically compare similarities between two or more different modalities other than text (i.e., mapping conceptually near and far between sets of images).

After demonstrating the distinctiveness of both near and far inspirational stimuli, this study then explored their impact on design outcomes. In this work, evaluations of midpoint and final design solutions were used to assess the performance of the designers. Specifically, external raters evaluated the feasibility, usefulness, and novelty of each design. Results showed that participants provided with no stimuli had significantly less feasible design solutions, while those provided with far stimuli had significantly less useful design solutions. From this perspective, participants provided with near stimuli benefitted more from the intervention than those provided with the far stimuli. This corroborates previous findings from the authors which suggests near inspirational stimuli may be more helpful than far inspirational stimuli [1]. In contrast to the far and no stimulus conditions, the designs in the near condition were not negatively affected in their feasibility or usefulness. Nonetheless, an important piece to the puzzle is still missing: the overall goodness of the stimuli themselves.

This work utilizes a measure to capture the overall “goodness” of solutions to assess design concepts. The motivation for this measure stems from the ambiguity in a common design metric prevalent throughout the engineering design field: quality. Many works utilize this metric when

assessing design artefacts, yet its holistic terminology and the variations in its definition, sometimes cloud the dimensions underlying its true meaning. For this reason, this work explores the use of a different measure. Literature review supports the concepts of feasibility and usefulness when considering the overall goodness of a design. Both are included in our newly defined *design innovation measure*. Novelty, while not as common of a sub-dimension for overall goodness, adds the element of uniqueness to the measure.

In this work, different forms of the innovation measure are presented, motivated, and analyzed. Correlations with quality corroborate the underlying dimensions and the specific formulation proposed in this work (i.e., Equation 1) [27,44]. These four specific formulations were considered to explore a general subset of variations on the sub-dimensions and understand the sensitivity of the formulation to these variations. Future work should further investigate the stability and robustness of the proposed formulation with additional datasets.

Participants' midpoint and final solutions, as well as the provided inspirational stimulus, were analyzed using the newly defined design innovation measure. A similar analysis was carried out as described previously, but this time looking at the change in innovation scores between final designs and midpoint designs. Again, no significant change existed between conditions (near, far, or no stimulus). This suggests that providing an adaptive stimulus, either semantically near or semantically far, does not improve the innovation potential of solutions. In contrast, the innovation score of the provided stimulus did impact designers' outcomes: when provided with an inspirational stimulus with a relatively higher innovation score, designers were more likely to produce a more innovative final design, regardless of inspirational stimulus condition.

Previous research has proposed a “sweet spot” for analogical stimuli, representing an analogy that lies somewhere in between the near and far fields and yields the most benefit for positive

design outcomes [5]. Yet, the stimuli in this experiment occupy the far ends of the spectrum, as opposed to this sweet spot. With the analytic nature of LSA, perhaps a more precise designation of this sweet spot can be identified in order to understand where between the near and far fields this sweet spot truly lies. Based on the present study, it may be important to not only consider the distance of the provided stimulus, but also its relative innovative potential.

While the stimulus conditions did not outperform the control condition on all dimensions of design metric outcomes, one explanation for this underperformance is the interruption endured during problem solving. Prior research is inconclusive regarding whether brief interruptions hurt or help problem solvers. For example, work by Gero et al. demonstrated that interruptions during design ideation are hurtful due to a cognitive shift from primary to secondary tasks [45]. Other works support this claim for problem solving in general, and have considered ways to mitigate the effects of disruptions [46–48]. Conversely, work by Sio et al. found that interruptions are beneficial [49]. Despite this, much of the previous research on the effects of design stimuli on design outcomes does not specifically study timing as a variable. Research from Tseng et al. found that stimuli are most helpful after the development of an open goal [15]. One thought is that the time when stimuli are typically provided during design studies may not occur during a period of deep problem solving (e.g., at the beginning of an experiment), and therefore do not cause this cognitive shift to occur. The timing of the stimulus intervention is not specifically studied in this work; however, it is an area in need of future investigation.

Another factor that may have impacted the inspirational stimuli conditions is team versus individual efforts. Providing example solutions halfway through problem solving may be analogous to two members on a team interacting or sharing ideas with each other at a set interval (i.e., independent work on the same design challenge with one opportunity to exchange current

ideas/solutions). Because of deficiencies in group dynamics, nominal teams (teams composed of individuals who do not collaborate during problem solving) have been shown to outperform teams in a variety of problem domains including brainstorming, conceptual design, configuration design, and verbalization tasks [50–53]. Under this theory, designers that did not receive an example solution should perform better because they did not collaborate with the computer team member (nominal teams). Those that did receive an example solution are hindered because of the interaction with the computer. To fully corroborate this theory, future work is needed to study this type of human-computer interaction of problem solving in “hybrid team” (human-computer) environments. The computational framework in this work provides promise for an effective design tool of the future. Forthcoming research can address how and when to intervene during design problem solving. More specifically, these open questions involve which modalities of interventions are best for designers, and when during the problem-solving process is most effective for them to be applied.

6. CONCLUSION

This work utilized Latent Semantic Analysis (LSA) to adaptively select relevant inspirational stimuli to aid designers during a cognitive study. Sixty designers were split into three conditions: two conditions that modulated the distance of the provided inspirational stimulus and a control condition in which no stimulus was provided. The stimuli were selected based on the LSA comparison between the current status of the designers’ output and a database of design solutions. One key contribution of this work is the adaptive determination of which stimulus to provide to a designer based on their current output of design activity. Results indicate that LSA is a viable technique to make interventions with inspirational design stimuli. Using a newly defined measure of design innovation, this work also investigates the impact of the inspirational stimuli intervention

and design output. The overall innovativeness of the provided stimuli significantly correlated with the overall innovativeness of the designers' final design solutions. In fact, the overall innovativeness of a stimulus had a greater impact on a designer's output than the relative distance of the stimulus. This highlights the need to provide stimuli to designers at specific distances relative to the solution space, while also assessing the innovative potential of the inspirational stimulus. While more work is needed to automate the process of providing designers with positively impactful inspirational stimuli, the real-time computational approach presented in this work is a critical step towards realizing this goal.

ACKNOWLEDGMENT

This work has been supported by the Air Force Office of Scientific Research (AFOSR) through grant FA9550-18-0088. The findings presented in this work represent the views of the authors and not necessarily those of the sponsors. The authors would like to thank Leah Chong and Daniel Clymer for their assistance rating design solutions and transcribing inspirational stimuli. This paper is based on preliminary work from the 2019 International Design Engineering and Technical Conferences [44].

REFERENCES

- [1] Goucher-Lambert, K., and Cagan, J., 2019, "Crowdsourcing Inspiration: Using Crowd Generated Inspirational Stimuli to Support Designer Ideation," *Des. Stud.*, **61**, pp. 1–29.
- [2] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples," *J. Mech. Des.*, **133**(8), p. 081004.
- [3] Linsey, J. S., Wood, K. L., and Markman, A. B., 2008, "Modality and Representation in Analogy," *AI EDAM*, pp. 85–100.
- [4] Moreno, D. P., Hernandez, A. a., Yang, M. C., Otto, K. N., Holtta-Otto, K., Linsey, J. S.,

- Wood, K. L., and Linden, A., 2014, “Fundamental Studies in Design-by-Analogy: A Focus on Domain-Knowledge Experts and Applications to Transactional Design Problems,” *Des. Stud.*, **35**(3), pp. 232–272.
- [5] 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,” *J. Mech. Des.*, **135**(2), p. 021007.
- [6] Linsey, J. S., Markman, a. B., and Wood, K. L., 2012, “Design by Analogy: A Study of the WordTree Method for Problem Re-Representation,” *J. Mech. Des.*, **134**(4), pp. 041009–041009.
- [7] Cagan, J., Kotovsky, K., and McComb, C., 2018, “Are Teams Teams? And Other Musings on Humpty Dumpty’s Use of Language In The Design World,” *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 1–3.
- [8] Gentner, D., 1983, “Structure-Mapping: A Theoretical Framework for Analogy,” *Cogn. Sci.*, **7**(2), pp. 155–170.
- [9] Forbus, K. D., Gentner, D., and Law, K., 1995, “MAC/FAC: A Model of Similarity-Based Retrieval,” *Cogn. Sci.*, **19**(2), pp. 141–205.
- [10] Krawczyk, D. C., McClelland, M. M., Donovan, C. M., Tillman, G. D., and Maguire, M. J., 2010, “An FMRI Investigation of Cognitive Stages in Reasoning by Analogy.,” *Brain Res.*, **1342**, pp. 63–73.
- [11] 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,” *Des. Stud.*, **60**, pp. 1–38.
- [12] Chan, J., Dow, S. P., and Schunn, C. D., 2015, “Do the Best Design Ideas (Really) Come from Conceptually Distant Sources of Inspiration?,” *Des. Stud.*, **36**(C), pp. 31–58.
- [13] Moss, J., Kotovsky, K., and Cagan, J., 2007, “The Influence of Open Goals on the Acquisition of Problem-Relevant Information.,” *J. Exp. Psychol. Learn. Mem. Cogn.*, **33**(5), pp. 876–891.
- [14] Moss, J., Kotovsky, K., and Cagan, J., 2011, “The Effect of Incidental Hints When Problems Are Suspended before, during, or after an Impasse.,” *J. Exp. Psychol. Learn. Mem. Cogn.*, **37**(1), pp. 140–148.

- [15] Tseng, I., Moss, J., Cagan, J., and Kotovsky, K., 2008, “The Role of Timing and Analogical Similarity in the Stimulation of Idea Generation in Design,” *Des. Stud.*, **29**(3), pp. 203–221.
- [16] Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., and Yaniv, I., 1995, “Demystification of Cognitive Insight: Opportunistic Assimilation and the Prepared-Mind Perspective,” *The Nature of Insight*.
- [17] Shi, F., Chen, L., Han, J., and Childs, P., 2017, “A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval,” *J. Mech. Des. Trans. ASME*, **139**(11), pp. 1–14.
- [18] Han, J., Shi, F., Chen, L., and Childs, P. R. N., 2018, “A Computational Tool for Creative Idea Generation Based on Analogical Reasoning and Ontology,” *Artif. Intell. Eng. Des. Anal. Manuf. AIEDAM*, **32**(4), pp. 462–477.
- [19] Han, J., Shi, F., Chen, L., and Childs, P. R. N., 2018, “The Combinator-a Computer-Based Tool for Creative Idea Generation Based on a Simulation Approach,” *Des. Sci.*, **4**, pp. 1–34.
- [20] Pennington, J., Socher, R., and Manning, C., 2014, “Glove: Global Vectors for Word Representation,” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.
- [21] Gilon, K., Ng, F. Y., Chan, J., Assaf, H. L., Kittur, A., and Shahaf, D., 2018, “Analogy Mining for Specific Design Needs,” *Conference on Human Factors in Computing Systems - Proceedings*.
- [22] Chan, J., Chee Chang, J., Hope, T., Shahaf, D., and Kittur, A., 2018, “SOLVENT: A Mixed Initiative System for Finding Analogies between Research Papers,” *Proc. ACM Hum.-Comput. Interact.*, pp. 1–21.
- [23] Gyory, J. T., Cagan, J., and Kotovsky, K., 2019, “Are You Better off Alone? Mitigating the Underperformance of Engineering Teams during Conceptual Design through Adaptive Process Management,” *Res. Eng. Des.*, **30**(1), pp. 85–102.
- [24] Dong, A., 2005, “The Latent Semantic Approach to Studying Design Team Communication,” *Des. Stud.*, **26**(5), pp. 445–461.
- [25] Dong, A., Hill, A. W., and Agogino, A. M., 2004, “A Document Analysis Method for Characterizing Design Team Performance,” *J. Mech. Des.*, **126**(3), pp. 378–385.
- [26] Fu, K., Cagan, J., Kotovsky, K., and Wood, K., 2013, “Discovering Structure in Design Databases Through Functional and Surface Based Mapping,” *J. Mech. Des.*, **135**(3), p.

031006.

- [27] Gyory, J. T., Goucher-Lambert, K., Kotovsky, K., and Cagan, J., 2019, "Exploring the Application of Network Analytics in Characterizing a Conceptual Design Space," *Proc. Des. Soc. Int. Conf. Eng. Des.*, **1**(1), pp. 1953–1962.
- [28] Landauer, T. K., Foltz, P. W., and Laham, D., 1998, "An Introduction to Latent Semantic Analysis," *Discourse Process*.
- [29] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003, "Metrics for Measuring Ideation Effectiveness," *Des. Stud.*, **24**(2), pp. 111–134.
- [30] Ahmed, F., Ramachandran, S. K., Fuge, M., Hunter, S., and Miller, S., 2019, "Interpreting Idea Maps: Pairwise Comparisons Reveal What Makes Ideas Novel," *J. Mech. Des. Trans. ASME*, **141**(2).
- [31] Hu, W.-L., and Reid, T., 2018, "The Effects of Designers' Contextual Experience on the Ideation Process and Design Outcomes," *J. Mech. Des.*, **140**(10), p. 101101.
- [32] Kennedy, C., 2013, "Two Sources of Subjectivity : Qualitative Assessment and Dimensional Uncertainty Two Sources of Subjectivity : Qualitative Assessment and Dimensional Uncertainty," *Inquiry*, **56**(2–3), pp. 258–277.
- [33] Dieter, G. E., and Schmidt, L. C., 2013, *Engineering Design*, McGraw-Hill, New York, NY.
- [34] Verhaegen, P.-A., Vandevenne, D., and Duflou, J. R., 2012, "Originality and Novelty: A Different Universe," *International Design Conference - Design 2012*, Dubrovnik - Croatia, pp. 1961–1966.
- [35] Nieto, M. J., and Santamaría, L., 2007, "The Importance of Diverse Collaborative Networks for the Novelty of Product Innovation," *Technovation*, **27**(6–7), pp. 367–377.
- [36] Stephen, G. G., Gavin, M., and Lynda, A.-S., 1995, "Assessing a Multidimensional Measure of Radical Technological Innovation," *42*(3), p. 232.
- [37] Miller, S. R., Bailey, B. P., and Kirlik, A., 2014, "Exploring the Utility of Bayesian Truth Serum for Assessing Design Knowledge.," *Human-Computer Interact.*, **29**(5–6), pp. 487–515.
- [38] Goucher-Lambert, K., and Cagan, J., 2017, "Using Crowdsourcing to Provide Analogies for Designer Ideation in a Cognitive Study," *International Conference on Engineering Design*, Vancouver, B.C, pp. 1–11.
- [39] Cicchetti, D. V., 1994, "Guidlines, Criteria, and Rules of Thumb for Evaluating Normed

and Standardized Assessment Instruments in Psychology," *Psychol. Assess.*, **6**(4), pp. 284–290.

- [40] Viswanathan, V. K., and Linsey, J. S., 2013, "Design Fixation and Its Mitigation: A Study on the Role of Expertise," *Journal of Mechanical Design*, p. 051008.
- [41] Wilson, J. O., Rosen, D., Nelson, B. a., and Yen, J., 2010, "The Effects of Biological Examples in Idea Generation," *Des. Stud.*, **31**(2), pp. 169–186.
- [42] Hofmann, T., 2017, "Probabilistic Latent Semantic Indexing," *ACM SIGIR Forum*, **51**(2), pp. 211–218.
- [43] Blei, D. M., Ng, A. Y., and Jordan, M. I., 2003, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, **3**, pp. 993–1022.
- [44] Goucher-Lambert, K., Gyory, J. T., Kotovsky, K., and Cagan, J., 2019, "Computationally Derived Adaptive Inspirational Stimuli for Real-Time Design Support During Concept Generation," *ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Anaheim, CA, pp. 1–10.
- [45] Gero, J. S., Jiang, H., Dobolyi, K., Bellows, B., and Smythwood, M., 2015, "How Do Interruptions During Designing Affect Design Cognition," *Design Computing and Cognition '14*, pp. 119–133.
- [46] Cades, D. M., Boehm-Davis, D. A., Trafton, J. G., and Monk, C. A., 2011, "Mitigating Disruptive Effects of Interruptions Through Training: What Needs to Be Practiced?," *J. Exp. Psychol. Appl.*, **17**(2), pp. 97–109.
- [47] Hess, S. M., and Detweiler, M. C., 1994, "Training to Reduce the Disruptive Effects of Interruptions," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 1173–1177.
- [48] Ratwani, R. M., and Trafton, J. G., 2011, "A Real-Time Eye Tracking System for Predicting and Preventing Postcompletion Errors," *Human-Computer Interact.*, **26**(3), pp. 205–245.
- [49] Sio, U. N., Kotovsky, K., and Cagan, J., 2017, "Interrupted: The Roles of Distributed Effort and Incubation in Preventing Fixation and Generating Problem Solutions," *Mem. Cogn.*, **45**(4), pp. 553–565.
- [50] Mullen, B., Johnson, C., and Salas, E., 1991, "Productivity Loss in Brainstorming Groups: A Meta-Analytic Integration Productivity Loss in Brainstorming Groups," *Basic Appl. Soc. Psych.*, **12**(1), pp. 3–23.

- [51] Gyory, J. T., Cagan, J., and Kotovsky, K., 2018, "Should Teams Collaborate During Conceptual Engineering Design?: An Experimental Study," *ASME International Design Engineering Technical Conference*, Quebec City, Canada.
- [52] Sio, U. N., Kotovsky, K., and Cagan, J., 2018, "Silence Is Golden: The Effect of Verbalization on Group Performance," *J. Exp. Psychol. Gen.*, **147**(6), pp. 939–944.
- [53] McComb, C., Cagan, J., and Kotovsky, K., 2017, "Optimizing Design Teams Based on Problem Properties: Computational Team Simulations and an Applied Empirical Test," *J. Mech. Des.*, **139**(4), p. 041101.

FIGURE CAPTION LIST

- Fig. 1 Outline of cognitive study
- Fig. 2 Example midpoint designs and respective provided stimuli for both the near and far condition
- Fig. 3 The meaning of near and far stimuli: two experimental conditions were defined using latent semantic analysis and the conditions were significantly distinct from one another (boxes show upper and lower quartiles)
- Fig. 4 The impact of inspirational stimuli on final designs. The left side of figures 4A and 4B conceptually illustrates the distance measured in the accompanying plots. A) Participants' final design concepts are more similar to the stimulus when provided a near stimulus; B) far inspirational stimuli led to more relative movement (i.e., distance) within the design space.
- Fig. 5 Mean difference between final (D_2) and intermediate (D_1) designs for three metrics, separated by experimental condition
- Fig. 6 The overall innovativeness of the provided inspirational stimuli is predictive of the final design outcome

TABLE CAPTION LIST

Table 1 Formulations of design innovation measure, *I*

Table 2 Correlations between quality and four formulations of the design innovation measure

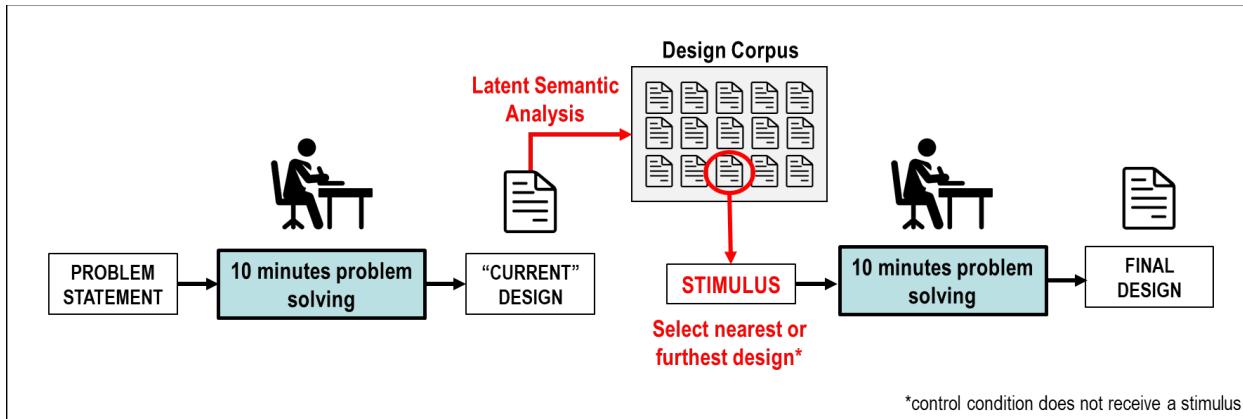


FIGURE 1: OUTLINE OF COGNITIVE STUDY

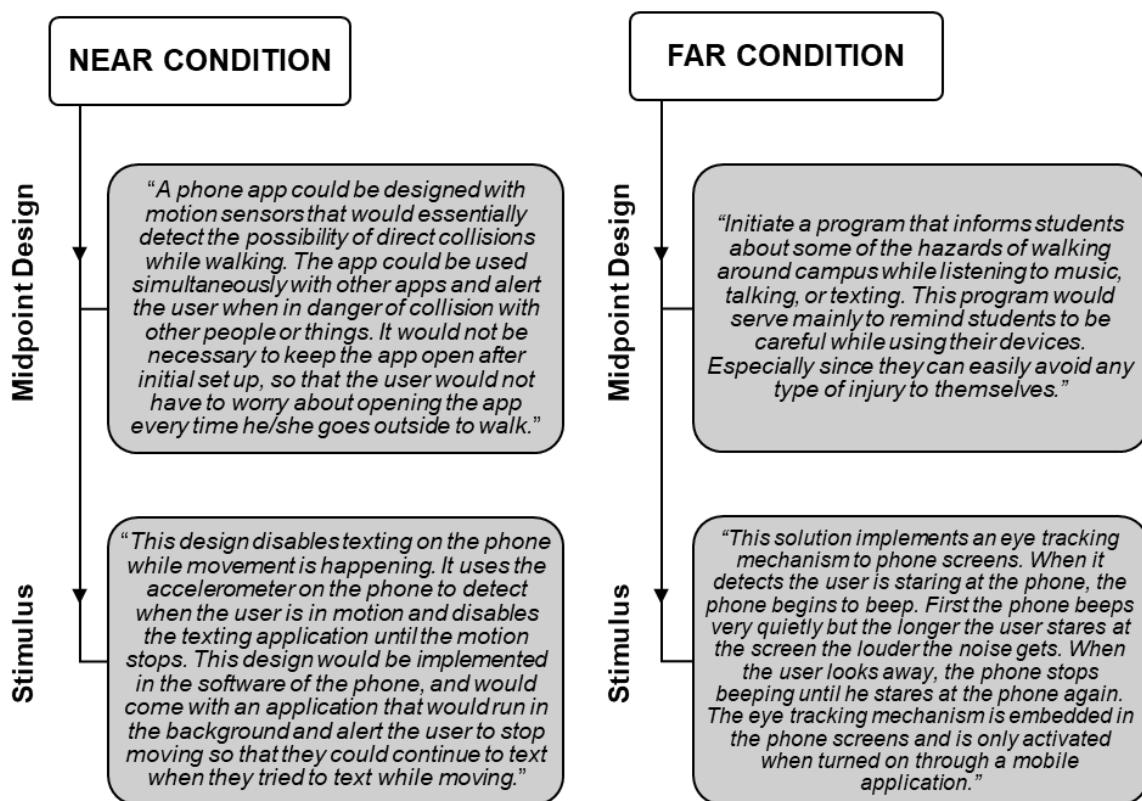


FIGURE 2: EXAMPLE MIDPOINT DESIGNS AND RESPECTIVE PROVIDED STIMULI FOR BOTH THE NEAR AND FAR CONDITION

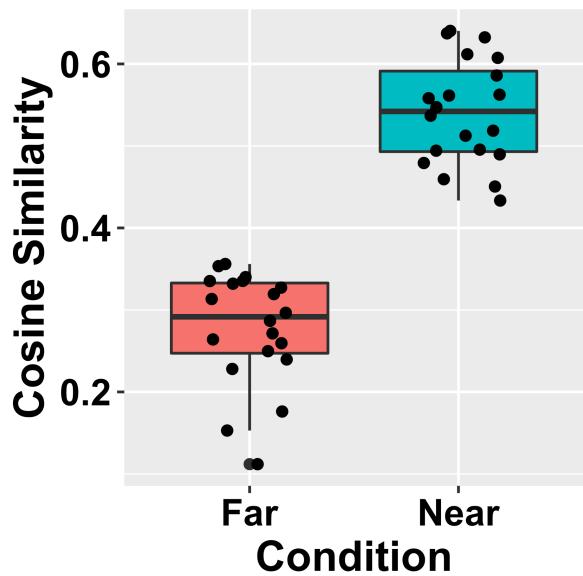


FIGURE 3: THE MEANING OF NEAR AND FAR STIMULI: TWO EXPERIMENTAL CONDITIONS WERE DEFINED USING LATENT SEMANTIC ANALYSIS AND THE CONDITIONS WERE SIGNIFICANTLY DISTINCT FROM ONE ANOTHER (BOXES SHOW UPPER AND LOWER QUARTILES)

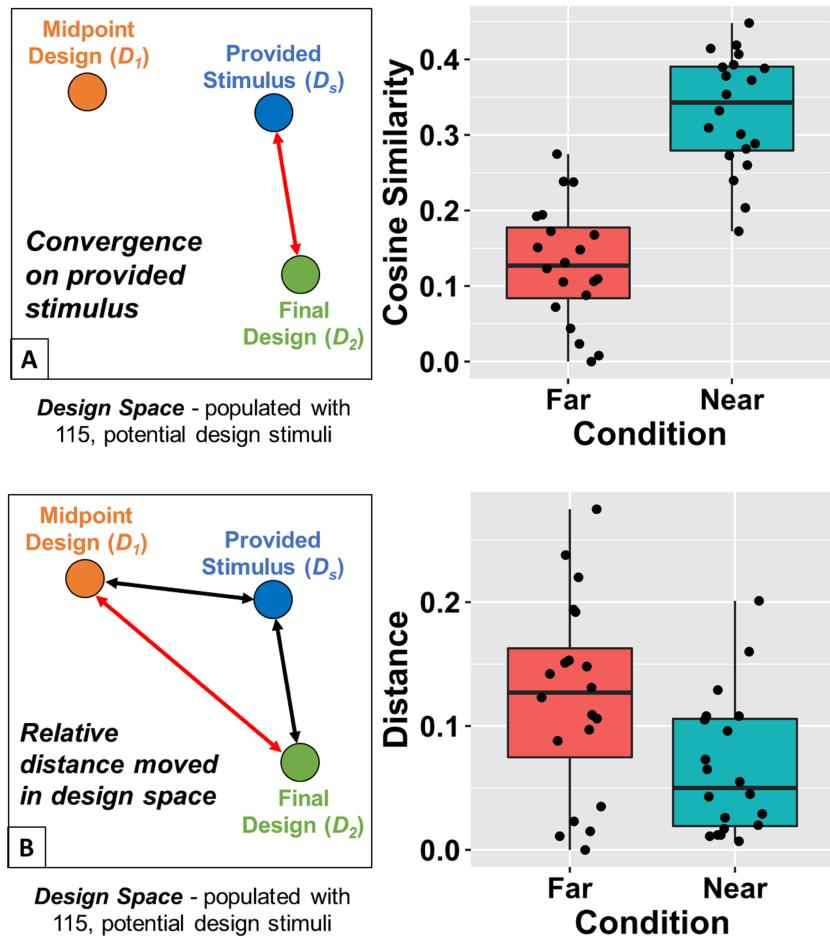


FIGURE 4: THE IMPACT OF INSPIRATIONAL STIMULI ON FINAL DESIGNS. THE LEFT SIDE OF FIGURES 4A AND 4B CONCEPTUALLY ILLUSTRATES THE DISTANCE MEASURED IN THE ACCOMPANYING PLOTS. A) PARTICIPANTS' FINAL DESIGN CONCEPTS ARE MORE SIMILAR TO THE STIMULUS WHEN PROVIDED A NEAR STIMULUS; B) FAR INSPIRATIONAL STIMULI LED TO MORE RELATIVE MOVEMENT (I.E., DISTANCE) WITHIN THE DESIGN SPACE.

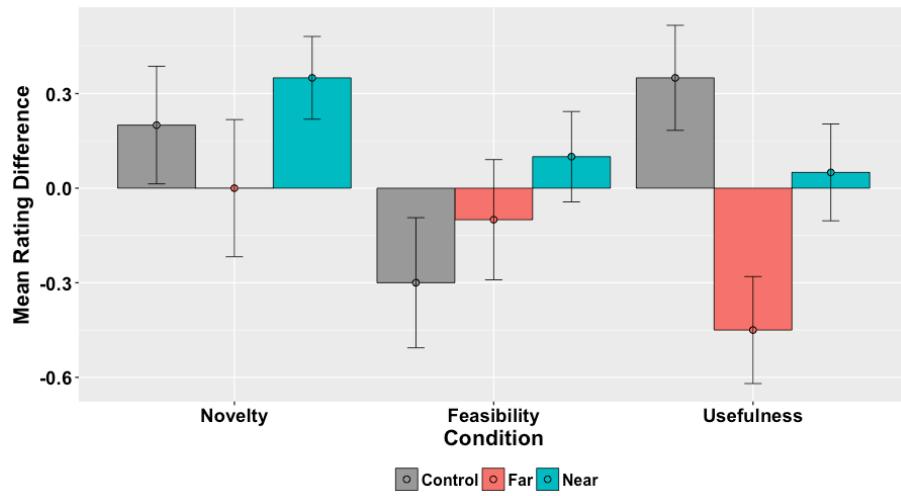


FIGURE 5: MEAN DIFFERENCE BETWEEN FINAL (D_2) AND INTERMEDIATE (D_1) DESIGNS FOR THREE METRICS, SEPARATED BY EXPERIMENTAL CONDITION

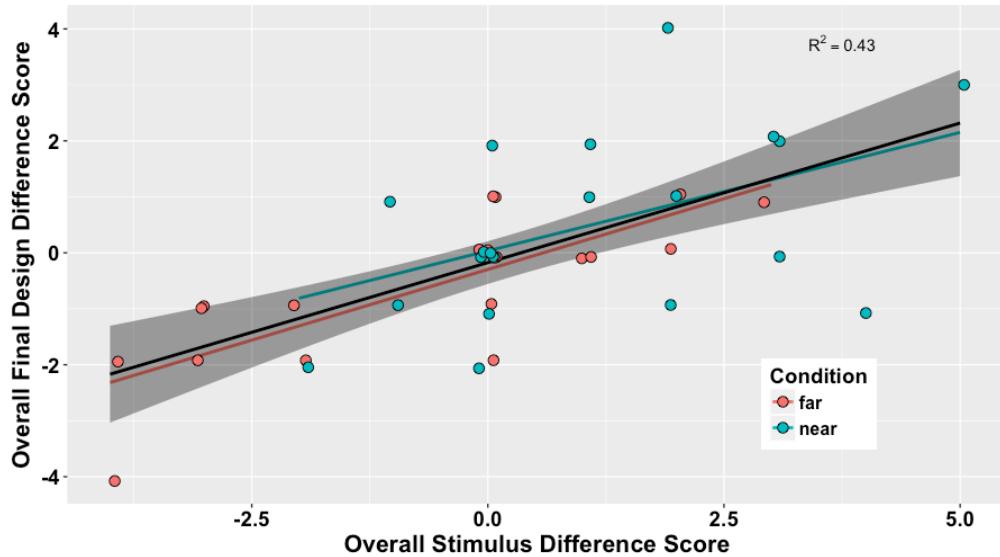


FIGURE 6: THE OVERALL INNOVATIVENESS OF THE PROVIDED INSPIRATIONAL STIMULI IS PREDICTIVE OF THE FINAL DESIGN OUTCOME

TABLE 1: FORMULATIONS OF DESIGN INNOVATION MEASURE, I

$I = (\mathbf{F} \cdot \mathbf{U}) + N$	(1)
$I = \mathbf{F} + \mathbf{U} + N$	(2)
$I = \mathbf{F} \cdot \mathbf{U} \cdot N$	(3)
$I = w_1\mathbf{F} + w_2\mathbf{U} + w_3N$	(4)

TABLE 2: CORRELATIONS BETWEEN QUALITY AND FOUR FORMULATIONS OF THE DESIGN INNOVATION MEASURE

Equation Problem \	(1)	(2)	(3)	(4)
Electricity	0.60	0.56	0.46	0.57
Phone	0.80	0.74	0.71	0.73
Joint	0.72	0.58	0.044	0.71
Surface	0.64	0.60	0.40	0.62