Do the best design ideas (really) come from conceptually distant sources of inspiration?

Joel Chan, Learning Research and Development Center, University of Pittsburgh, LRDC Room 823, 3939 O'Hara St, Pittsburgh, PA 15260, USA
Steven P. Dow, Human–Computer-Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Christian D. Schunn, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA

Design ideas often come from sources of inspiration (e.g., analogous designs, prior experiences). In this paper, we test the popular but unevenly supported hypothesis that conceptually distant sources of inspiration provide the best insights for creative production. Through text analysis of hundreds of design concepts across a dozen different design challenges on a Web-based innovation platform that tracks connections to sources of inspiration, we find that citing sources is associated with greater creativity of ideas, but conceptually closer rather than farther sources appear more beneficial. This inverse relationship between conceptual distance and design creativity is robust across different design problems on the platform. In light of these findings, we revisit theories of design inspiration and creative cognition.

Keywords: innovation, design cognition, creative design, conceptual design, sources of inspiration

Where do creative design ideas come from? Cognitive scientists have discovered that people inevitably build new ideas from their prior knowledge and experiences (Marsh, Ward, & Landau, 1999; Ward, 1994). While these prior experiences can serve as sources of inspiration (Eckert & Stacey, 1998) and drive sustained creation of ideas that are both new and have high potential for impact (Hargadon & Sutton, 1997; Helms, Vattam, & Goel, 2009), they can also lead designers astray: for instance, designers sometimes incorporate undesirable features from existing solutions (Jansson & Smith, 1991; Linsey et al., 2010), and prior knowledge can make it difficult to think of alternative approaches (German & Barrett, 2005; Wiley, 1998). This raises the question: what features of potential inspirational sources can predict their value (and/or potential harmful effects)? In this paper, we examine how the conceptual distance of sources relates to their inspirational value.

Corresponding author:
Joel Chan
joc59@pitt.edu
joelchuc@cs.cmu.edu

www.elsevier.com/locate/destud
0142-694X *Design Studies* ■■ (2014) ■■–■■
http://dx.doi.org/10.1016/j.destud.2014.08.001

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

1 Background

1.1 Research base

What do we mean by conceptual distance? Consider the problem of e-waste accumulation: the world generates 20–50 million metric tons of e-waste every year, yielding environmentally hazardous additions to landfills. A designer might approach this problem by building on **near** sources like smaller-scale electronics reuse/recycle efforts, or by drawing inspiration from a **far** source like edible food packaging technology (e.g., to design re-usable electronics parts). What are the relative benefits of different levels of source conceptual distance along a continuum from near to far?

Many authors, principally those studying the role of analogy in creative problem solving, have proposed that conceptually far sources — structurally similar ideas with many surface (or object) dissimilarities — are the best sources of inspiration for creative breakthroughs (Gentner & Markman, 1997; Holyoak & Thagard, 1996; Poze, 1983; Ward, 1998). This proposal — here called the Conceptual Leap Hypothesis — is consistent with many anecdotal accounts of creative breakthroughs, from Kekule's discovery of the structure of benzene by visual analogy to a snake biting its tail (Findlay, 1965), to George Mestral's invention of Velcro by analogy to burdock root seeds (Freeman & Golden, 1997), to more recent case studies (Enkel & Gassmann, 2010; Kalogerakis, Lu, & Herstatt, 2010).

However, empirical support for this proposal is mixed. Some studies have shown an advantage of far over near sources for novelty, quality, and flexibility of ideation (Chan et al., 2011; Chiu & Shu, 2012; Dahl & Moreau, 2002; Gonçalves, Cardoso, & Badke-Schaub, 2013; Hender, Dean, Rodgers, & Jay, 2002); but, some *in vivo* studies of creative cognition have not found strong connections between far sources and creative mental leaps (Chan & Schunn, 2014; Dunbar, 1997), and other experiments have demonstrated equivalent benefits of far and near sources (Enkel & Gassmann, 2010; Malaga, 2000). Relatedly, Tseng, Moss, Cagan, and Kotovsky (2008) showed that far sources were more impactful after ideation had already begun (vs. before ideation), providing more functionally distinct ideas than near or control, but both far and near sources led to similar levels of novelty. Similarly, Wilson, Rosen, Nelson, and Yen (2010) showed no advantage of far over near sources for novelty of ideas (although near but not far sources decreased variety of ideas). Fu et al. (2013) even found that far sources led to lower novelty and quality of ideas than near sources. Thus, more empirical work is needed to determine whether the Conceptual Leap Hypothesis is well supported. Further, Fu et al. (2013) argue there is an inverted U-shape function in which moderate distance is best, suggesting the importance of conceptualizing and measuring distance along a continuum.

1.2 Impetus for the current work

Key methodological shortcomings in prior work further motivate more and better empirical work. Prior studies may be too short (typically 30 min to 1 h) to convert far sources into viable concepts. To successfully use far sources, designers must spend considerable cognitive effort to ignore irrelevant surface details, attend to potentially insightful structural similarities, and adapt the source to the target context. Additionally, many far sources may yield shallow or unusable inferences (e.g., due to non-alignable differences in structural or surface features; Perkins, 1997); thus, designers might have to sift through many samples of far sources to find 'hidden gems.' These higher processing costs for far sources might partially explain why some studies show a negative impact of far sources on the number of ideas generated (Chan et al., 2011; Hender et al., 2002). In the context of a short task, these processing costs might take up valuable time and resources that could be used for other important aspects of ideation (e.g., iteration, idea selection); in contrast, in real-world design contexts, designers typically have days, weeks or even months (not an hour) to consider and process far sources.

A second issue is a lack of statistical power. Most existing experimental studies have $N \leq 12$ per treatment cell (Chiu & Shu, 2012; Hender et al., 2002; Malaga, 2000); only four studies had $N \geq 18$ (Chan et al., 2011; Fu et al., 2013; Gonçalves et al., 2013; Tseng et al., 2008), and they are evenly split in support/opposition for the benefits of far sources. Among the few correlational studies, only Dahl and Moreau (2002) had a well powered study design in this regard, with 119 participants and a reasonable range of conceptual distance. Enkel and Gassmann (2010) only examined 25 cases, all of which were cases of cross-industry transfer (thus restricting the range of conceptual distance being considered). This lack of statistical power may have led to a proliferation of false negatives (potentially exacerbated by small or potentially zero effects at short time scales), but possibly also severely overestimated effect sizes or false positives (Button et al., 2013); more adequately powered studies are needed for more precise estimates of the effects of conceptual distance.

A final methodological issue is problem variation. Many experimental studies focused on a single design problem. The inconsistent outcomes in these studies may be partially due to some design problems having unique characteristics, e.g., coincidentally having good solutions that overlap with concepts in far sources. Indeed, Chiu and Shu (2012), who examined multiple design problems, observed inconsistent effects across problems. Other investigations of design stimuli have also observed problem variation for effects (Goldschmidt & Smolkov, 2006; Liikkanen & Perttula, 2008).

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

This paper contributes to theories of design inspiration by 1) reporting the results of a study that addresses these methodological issues to yield clearer evidence, and 2) (to foreshadow our results) re-examining theories of design inspiration and conceptual distance in light of accumulating preponderance of evidence *against* the Conceptual Leap Hypothesis.

2 Methods

2.1 Overview of research context

The current work is conducted in the context of OpenIDEO (www.openideo.com), a Web-based crowd-sourced innovation platform that addresses a range of social and environmental problems (e.g., managing e-waste, increasing accessibility in elections). The OpenIDEO designers, with expertise in design processes, guide contributors to the platform through a structured design process to produce concepts that are ultimately implemented for real-world impact ('Impact Stories,' n.d.). For this study, we focus on three crucial early stages in the process: first, in the *inspiration* phase (lasting between 1.5 and 4 weeks, $M = 3.1$), contributors post *inspirations* (e.g., descriptions of solutions to analogous problems and case studies of stakeholders), which help to define the problem space and identify promising solution approaches; then, in the *concepting* phase (lasting the next 2–6 weeks, $m = 3.4$), contributors post *concepts*, i.e., specific solutions to the problem. Figure 1 shows an example concept; it is representative of the typical length and level of detail in concepts, i.e., ~150 words on average, more detail than one or two words/sentences/sketches, but less detail than a full-fledged design report/presentation or patent application. Finally, a subset of these concepts is *shortlisted* by an expert panel (composed of the OpenIDEO designers and a set of domain experts/stakeholders) for further refinement, based on their creative potential. In later stages, these concepts are refined and evaluated in more detail, and then a subset of them is selected for implementation. We focus on the first three stages given our focus on creative *ideation* (the later stages involve many other design processes, such as prototyping).

The OpenIDEO platform has many desirable properties as a research context for our work, including the existence of multiple design problems, thousands of concepts and inspirations, substantive written descriptions of ideas to enable efficient text-based analyses, and records of feedback received for each idea, another critical factor in design success. A central property for our research question is the explicit nature of sources of inspiration in the OpenIDEO workflow. The site encourages contributors to build on others' ideas. Importantly, when posting concepts or inspirations, contributors are prompted to cite any concepts or inspirations that serve as sources of inspiration for their idea. Also, when browsing other concepts/inspirations, they are able to also see concepts/inspirations the given concept/inspiration 'built upon' (i.e., cited as explicit sources of inspiration; see Figure 2). This culture

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

E-trash into real cash

Companies can end up with left-over electronics and components for electronics, imagine if there was a marketplace for them to sell their scrap, trash, and left-over chemicals to other companies that need it.

Example Use cases:

1: Big Corp makes 50,000 widgets that need ingredient A in the casing. Unfortunately, the widgets are discontinued and Big Corp is left with mountains of ingredient A that they don't foresee using the future. They are about to throw it all away since they need the space in their warehouse when Big Corp goes to E-trash.com and finds Fancy Corp who just decided to make 100,000 gizmos that really need ingredient A. E-trash facilitates the transaction and mountains of ingredient A don't go to the landfill!

2: Big Corp has thousands of version 1 doodads that they used for the last couple of years but now they need new version 2. They need to get rid of it quickly and so they go to E-trash.com and put it up to find out that Fancy Corp really needs doodads and version 1 works perfectly! Transaction made! Alternatively, version 1 just isn't applicable anymore but ingredient B in it could be very valuable so through E-trash.com they find a recycler who specializes in extracting ingredient B from old electronics and then selling it to other companies.

Description: It would be an on-line marketplace for businesses to find business buyers for their large quantities of e-waste. Sellers could post, either publicly or to select partners, what "waste" they have available and then buyers could bid on the "waste" that they could actually use. Lots of e-products and electronic components can be re-used and re-purposed. This would provide a method for companies to make money off of their waste and to find necessary products and components at a discount. This idea is in large part inspired by the company recyclematch.com. They focus more on traditional manufacturing components.

How does your concept safeguard human health and protect our environment?

It helps to prevent companies from disposing of large quantities of e-waste that other companies could really use.

Where does your concept fit into the lifecycle of electronic devices?

It fits in at the end/beginning of the lifecycle as one business loses the need for the e-waste components or end products that could serve as a foundation point for another company's products.

What steps could be taken today to start implementing your concept?

Encourage existing b2b waste management companies like recyclematch to pursue this by providing the business case of how much money is in working with e-waste especially as the necessary raw materials get harder to find.

What kinds of resources will be needed to fully implement and scale your concept?

Would need an on-line marketplace or perhaps a grant could be awarded to recyclematch or some other player in business to business (b2b) waste management who would already have the necessary connections and framework of a marketplace that could be build upon.

Figure 1 Example concept illustrating the typical amount of detail per concept

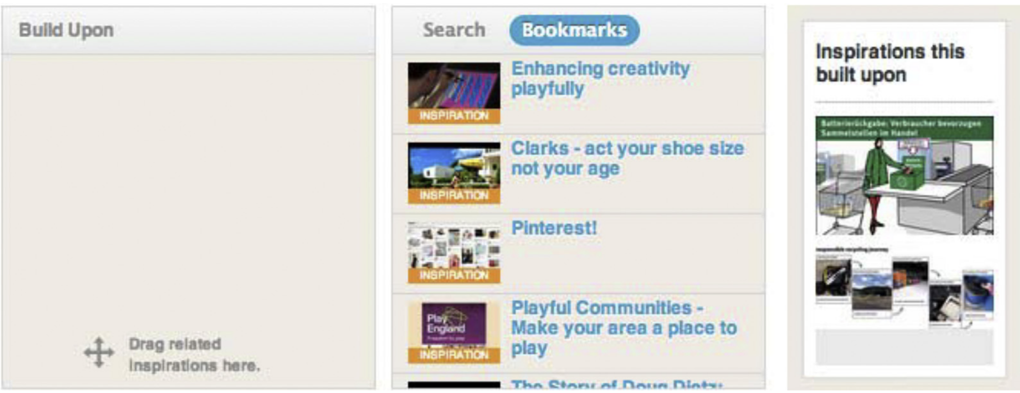

Figure 2 Depiction of OpenIDEO citation workflow. When posting concepts/inspirations, users are prompted to cite concepts/inspirations they 'build upon' by dragging bookmarked concepts/inspirations (middle panel) to the citation area (left panel). Users can also search for related concepts/inspirations at this step (middle panel). These cited sources then show up as metadata for the concept/inspiration (right panel)

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

of citing sources is particularly advantageous, given that people generally forget to monitor or cite their sources of inspiration (Brown & Murphy, 1989; Marsh, Landau, & Hicks, 1997), and our goal is to study the effects of source use. While users might still forget to cite sources, these platform features help ensure higher rates of source monitoring than other naturalistic ideation contexts. We note that this operationalization of sources as self-identified citations precludes consideration of implicit stimulation; however, the Conceptual Leap Hypothesis may be more applicable to conscious inspiration processes (e.g., analogy, for which conscious processing is arguably an important defining feature; Schunn & Dunbar, 1996).

2.2 Sample and initial data collection

The full dataset for this study consists of 2341 concepts posted for 12 completed challenges by 1190 unique contributors, citing 4557 unique inspirations; 241 (10%) of these concepts are shortlisted for further refinement. See Table 2 for a description of the 12 challenges (with some basic metadata on each challenge). Figure 3 shows the full-text design brief for two challenges.

How can we manage e-waste & discarded electronics to safeguard human health & protect our environment?

Ever wondered what happens to your outmoded cell phone when you replace it with the latest model? Or where a battery goes when you toss it in the trash? Around the world, end-of-life electronics that are waste, also known as e-waste, present a significant challenge for our environment and our health. Together with Brazilian bank Itaú Unibanco, the U.S. Department of State, and the U.S. Environmental Protection Agency, we're asking the OpenIDEO community to help us find ways to manage e-waste to better safeguard human health and protect our environment.According to the Consumer Electronics Association, in 2012 global spending on electronics is expected to surpass US$1 trillion. As use of these electronics increases around the world, the question of how to properly manage them when consumers are finished with them becomes more urgent. Unfortunately, not enough of our electronics are re-used, recycled or refurbished. Too many of them end up directly in landfills or recovered in an unsafe manner. According to the UN Environmental Programme, some 20 to 50 million metric tons of e-waste are generated worldwide every year, with mobile phones and televisions contributing 10 million tons per year by 20152. E-waste presents complex issues with many factors to consider, one of them being the environmental impact of hazardous substances and toxic chemicals – including lead, nickel, cadmium and mercury. The great news is that many of the materials used in consumer electronics can be recycled or refurbished to be used in other electronics. The opportunity is to find better ways to manage our used and end-of-life electronics and avoid them ending up in landfills.

How might we increase the number of registered bone marrow donors to help save more lives?

OpenIDEO has partnered with the Haas Center for Public Service at Stanford University to explore new ideas for encouraging bone marrow donation worldwide. Together we're asking you, the OpenIDEO community, to help us find ways to expand the global network of potential bone marrow donors and support people who are battling leukemia and other blood cancers. Bone marrow transplants are one type of treatment for leukemia and other blood or bone marrow cancers. This OpenIDEO challenge will complement the efforts of 100K Cheeks, a Stanford-based advocacy group dedicated to increasing the number of people enrolled in bone marrow registries worldwide. Certain populations are dramatically under-represented in existing bone marrow registries. For example, the match rate within the South Asian* demographic is critically low—with a 1 in 20,000 chance for a potential recipient to find a match. For more information about bone marrow donation (including the process, myths, and facts), visit BeTheMatch.org. If you're interested in becoming a donor, you can look up the registry in your country here.

Figure 3 Full-text of challenge briefs from two OpenIDEO challenges

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

With administrator permission, we downloaded all inspirations and concepts (which exist as individual webpages) and used an HTML parser to extract the following data and metadata:

1) Concept/inspiration author (who posted the concept/inspiration)
2) Number of comments (before the refinement phase)
3) Shortlist status (yes/no)
4) List of cited sources of inspiration
5) Full-text of concept/inspiration

Not all concepts cited inspirations as sources. Of the 2341 concepts, 707 (posted by 357 authors) cited at least one inspiration, collectively citing 2245 unique inspirations. 110 of these concepts (∼16%) were shortlisted (see Table 1 for a breakdown by challenge). This set of 707 concepts is the primary sample for this study; the others serve as a contrast to examine the value of explicit building at all on prior sources, and to aid in interpretation of any negative or positive effects of variations in distance. Because we only collected publicly available data, we do not have complete information on the expertise of all contributors: however, based on their public profiles on OpenIDEO, at least 1/3 of the authors in this sample are professionals in design-related disciplines (e.g., user experience/interaction design, communication design, architecture, product/industrial design, entrepreneurs and social innovators, etc.) and/or domain experts or stakeholders (e.g., urban development researcher

Table 1 Descriptions and number of posts for OpenIDEO challenges in final analysis sample

Name/description	*# of Inspirations*	*# of Concepts* (shortlisted)
How might we increase the number of registered bone marrow donors to help save more lives?	186	71 (*7*)
How might we inspire and enable communities to take more initiative in making their local environments better?	160	44 (*11*)
How can we manage e-waste & discarded electronics to safeguard human health & protect our environment?	60	26 (*8*)
How might we better connect food production and consumption?	266	147 (*10*)
How can technology help people working to uphold human rights in the face of unlawful detention?	248	62 (*7*)
How might we identify and celebrate businesses that innovate for world benefit and inspire other companies to do the same?	122	24 (*13*)
How might we use social business to improve health in low-income communities?	131	46 (*11*)
How might we increase social impact with OpenIDEO over the next year?	67	40 (*12*)
How might we restore vibrancy in cities and regions facing economic decline?	558	119 (*13*)
How might we design an accessible election experience for everyone?	241	47 (*8*)
How might we support web entrepreneurs in launching and growing sustainable global businesses?	88	49 (*7*)
How can we equip young people with the skills, information and opportunities to succeed in the world of work?	118	32 (*3*)

contributing to the vibrant-cities challenge, education policy researcher contributing to the youth-employment challenge, medical professional contributing to the bone-marrow challenge). Collectively, these authors accounted for approximately half of the 707 concepts in this study.

We analyze the impact of the distance of inspirations (and not cited concepts) given our focus on ideation processes during 'original' or non-routine design, where designers often start with a problem and only 'inspirations' (e.g., information about the problem or potentially related designs) rather than routine design (e.g., configuration or parametric design), where designers might be modifying or iterating on existing solutions rather than generating novel ones (Chakrabarti, 2006; Dym, 1994; Gero, 2000; Ullman, 2002). The Conceptual Leap Hypothesis maps most clearly to non-routine design.

2.3 Measures

2.3.1 Creativity of concepts

We operationalize concept creativity as whether a concept gets shortlisted. Shortlisting is done by a panel of expert judges, including the original challenge sponsors, who have spent significant time searching for and learning about existing approaches, and the OpenIDEO designers, who are experts in the general domain of creative design, and who have spent considerable time upfront with challenge sponsors learning about and defining the problem space for each challenge.

An expert panel is widely considered a 'gold standard' for measuring the creativity of ideas (Amabile, 1982; Baer & McKool, 2009; Brown, 1989; Sawyer, 2012). Further, we know from conversations with the OpenIDEO team that the panel's judgments combines consideration of both novelty and usefulness/appropriateness (here operationalized as potential for impact; A. Jablow, personal communication, May 1, 2014), the standard definition of creativity (Sawyer, 2012). Since OpenIDEO challenges are novel and unsolved, successful concepts are different from (and, perhaps more importantly, significantly better than) existing unsatisfactory solutions. We use shortlist (rather than win status) given our focus on the ideation phase in design (vs. convergence/refinement, which happens after concepts are shortlisted, and can strongly influence which shortlisted concepts get selected as 'winners' for implementation).

2.3.2 Conceptual distance

2.3.2.1 Measurement approach. Measuring conceptual distance is a major methodological challenge, especially when studying large samples of ideation processes (e.g., many designs across many design problems). The complex and multifaceted nature of typical design problems can make it difficult to distinguish 'within' and 'between' domain sources in a consistent and principled manner. Further, using only a binary scale risks losing variance information that could be critical for converging on a more precise understanding of the effects of conceptual distance (e.g., curvilinear effects across the continuum of distance). Continuous distance measures are an attractive alternative, but can be extremely costly to obtain at this scale, especially for naturalistic sources (e.g., relatively developed text descriptions vs. simple sketches or one-to-two sentence descriptions). Human raters may suffer from high levels of fatigue, resulting in poor reliability or drift of standards.

We address this methodological challenge with probabilistic topic modeling (Blei, 2012; Steyvers & Griffiths, 2007), a major computational approach for understanding large collections of unstructured text. They are similar to other unsupervised machine learning methods — e.g., *K*-means clustering, and Latent Semantic Analysis (Deerwester, Dumais, Furnas, & Landauer, 1990) — but distinct in that they emphasize human understanding of not just the relationship between documents in a collection, but the ‘reasons’ for the hypothesized relationships (e.g., the ‘meaning’ of particular dimensions of variation), largely because the algorithms underlying these models tend to produce dimensions in terms of clusters of tightly co-occurring words. Thus, they have been used most prominently in applications where understanding of a corpus, not just information retrieval performance, is a high priority goal, e.g., knowledge discovery and information retrieval in repositories of scientific papers (Griffiths & Steyvers, 2004), describing the structure and evolution of scientific fields (Blei & Lafferty, 2006, 2007), and discovering topical dynamics in social media use (Schwartz et al., 2013).

We use Latent Dirichlet Allocation (LDA; Blei, Ng, Jordan, & Lafferty, 2003), the simplest topic model. LDA assumes that documents are composed of a mixture of latent ‘topics’ (occurring with different ‘weights’ in the mixture), which in turn generate the words in the documents. LDA defines topics as probability distributions over words: for example, a ‘genetics’ topic can be thought of as a probability distribution over the words {phenotype, population, transcription, cameras, quarterbacks}, such that words closely related to the topic {phenotype, population, transcription} have a high probability in that topic, and words not closely related to the topic {cameras, quarterbacks} have a very low probability. Using Bayesian statistical learning algorithms, LDA infers the latent topical structure of the corpus from the co-occurrence patterns of words across documents. This topical structure includes 1) the topics in the corpus, i.e., the sets of probability distributions over words, and 2) the topic mixtures for each document, i.e., a vector of weights for each of the corpus topics for that document. We can derive conceptual *similarity* between any pair of documents by computing the cosine between their topic-weight vectors. In essence, documents that share dominant topics in similar relative proportions are the most similar.

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

Here, we used the open-source MAchine Learning for LanguagE Toolkit (MALLET; McCallum, 2002) to train an LDA model with 400 topics for all documents in the full dataset, i.e., 2341 concepts, 4557 inspirations, and 12 challenge briefs (6910 total documents). Additional technical details on the model-building procedure are available in Appendix A. Resulting cosines between inspirations and the challenge brief ranged from 0.01 to 0.91 ($M = 0.21$, $SD = 0.18$), a fairly typical range for large-scale information retrieval applications (Jessup & Martin, 2001).

2.3.2.2 *Validation*. Since we use LDA's measures of conceptual distance as a *substitute* for human judgments, we validate the adequacy of our topic model using measures of fit with human similarity judgments on a subset of the data by trained human raters.

Five trained raters used a Likert-type scale to rate 199 inspirations from one OpenIDEO challenge for similarity to their challenge brief, from 1 (very dissimilar) to 6 (extremely similar). Raters were given the intuition that the rating would approximately track the proportion of 'topical overlap' between each inspiration and the challenge brief, or the extent to which they are 'about the same thing.' The design challenge context was explicitly deemphasized, so as to reduce the influence of individual differences in perceptions of the 'relevance' of sources of inspiration. Thus, the raters were instructed to treat all the documents as 'documents' (e.g., an article about some topics, vs. 'problem solution') and consciously avoid judging the 'value' of the inspirations, simply focusing on semantic similarity. Raters listed major topics in the challenge brief and evaluated each inspiration against those major topics. To ensure internal consistency, the raters also sorted the inspirations by similarity after every 15–20 judgments. They then inspected the rank ordering and composition of inspirations at each point in the scale, and made adjustments if necessary (e.g., if an inspiration previously rated as '1' now, in light of newly encountered inspirations, seemed more like a '2' or '3'). Although the task was difficult, the mean ratings across raters had an acceptable aggregate consistency intra-class correlation coefficient (ICC(2,5)) of 0.74 (mean inter-coder correlation = 0.36). LDA cosines correlated highly, at $r = 0.51$, 95% CI = [0.40, 0.60], with the continuous human similarity judgments (see Figure 4A). We note that this correlation is better than the highest correlation between human raters ($r = 0.48$), reinforcing the value of automatic coding methods for this difficult task.

For comparability with prior work, we also measure fit with binary (within- vs. between-domain) distance ratings. Two raters also classified 345 inspirations from a different challenge as either within- or between-domain. Raters first collaboratively defined the problem domain, focusing on the question, 'What is the problem to be solved?' before rating inspirations. Within-domain inspirations were information about the problem (e.g., stakeholders, constraints)

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

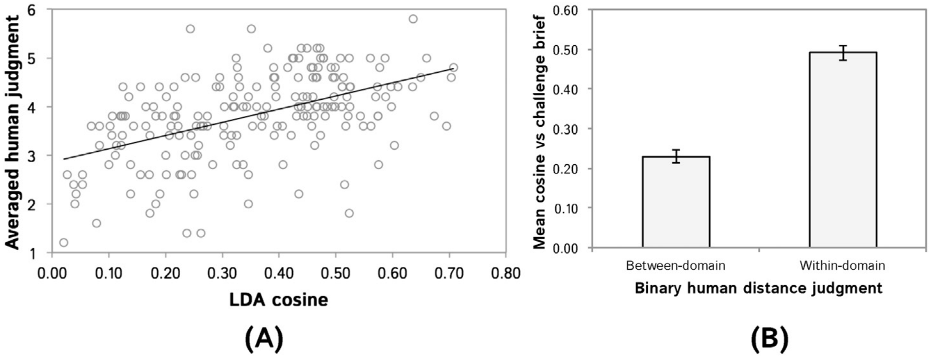

Figure 4 (A) Scatterplot of LDA cosines vs. averaged human continuous similarity judgments for inspirations in the e-waste challenge. (B). Mean cosine against the challenge brief for within- vs. between-domain inspirations

and existing prior solutions for very similar problems, while between-domain inspirations were information/solutions for analogous or different problems. Reliability for this measure was acceptable, with an overall average kappa of 0.78 (89% agreement). All disagreements were resolved by discussion. Similar to the continuous similarity judgments, the point biserial correlation between the LDA-derived cosine and the binary judgments was also high, at 0.50, 95% CI = [0.42, 0.58]. The mean cosine to the challenge brief was also higher for within-domain ($M = 0.49$, SD = 0.25, $N = 181$) vs. between-domain inspirations ($M = 0.23$, SD = 0.20, $N = 164$), $d = 1.16$, 95% CI = [1.13, 1.19] (see Figure 4B), further validating the LDA approach to measuring distance. Figure 5 shows examples of a near and far inspiration (from the e-waste challenge), along with the top 3 LDA topics (represented by the top 5 words for that latent topic), computed cosine vs. its challenge brief, and human similarity rating. The top 3 topics for the challenge brief are {waste, e, recycling, electronics, electronic}, {waste, materials, recycling, recycled, material}, and {devices, electronics, electronic, device, products}, distinguishing e-waste, general recycling, and electronics products topics. These examples illustrate how LDA is able to effectively extract the latent topical mixture of the inspirations from their text (inspirations with media also include textual descriptions of the media, mitigating concerns about loss of semantic information due to using only text as input to LDA) and also capture intuitions about variations in conceptual distance among inspirations: a document about different ways of assigning value to possessions is intuitively conceptually more distant from the domain of e-waste than a document about a prior effort to address e-waste.

The near and far examples depicted in Figure 5 also represent the range of conceptual distance measured in this dataset, with the near inspiration's cosine of 0.64 representing approximately the 90th percentile of similarity to the challenge domain, and the far inspiration's cosine of 0.01 representing approximately the 10th percentile of similarity to the challenge domain. Thus, the range of conceptual distance of inspirations in this data spans approximately

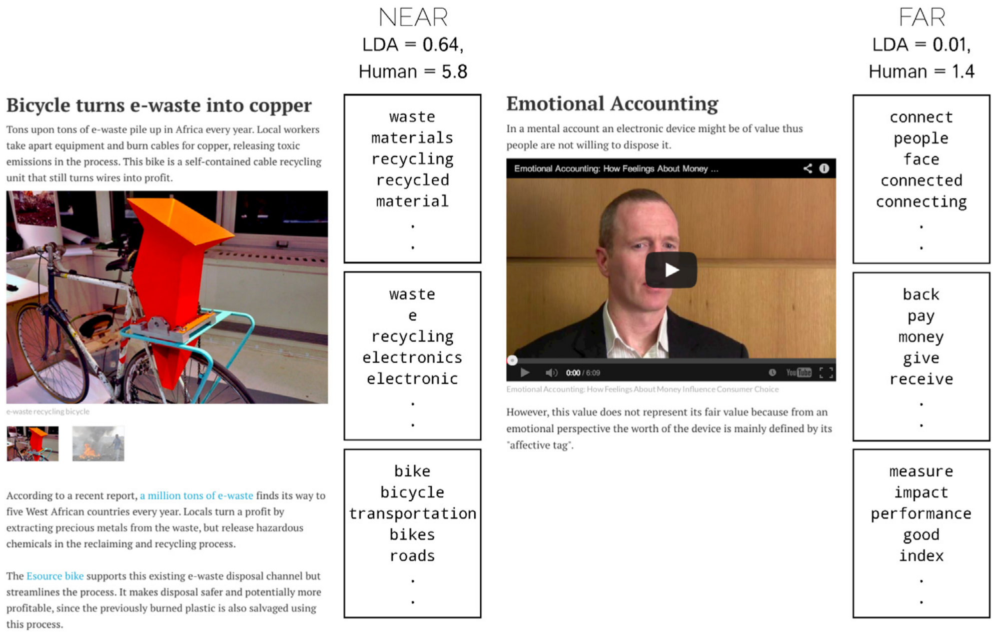

Figure 5 Topics found by LDA within examples of near and far inspirations for the e-waste challenge

from sources that are very clearly within the domain (e.g., an actual solution for the problem of electronic waste involving recycling of materials) to sources that are quite distant, but not obviously random (e.g., an observation of how people assign emotional value to relationships and artifacts). This range most likely excludes the 'too far' example designs studied in Fu et al. (2013) or the 'opposite stimuli' used in Chiu and Shu (2012).

2.3.2.3 *Final distance measures*. The challenge briefs varied in length and specificity across challenges, as did mean raw cosines for inspirations. But, these differences in mean similarity were much larger, $d = 1.90$, 95% CI $= [1.85-1.92]$ (for 80 inspirations from 4 challenges with maximally different mean cosines), than for human similarity judgments (coded separately but with the same methodology as before), $d = 0.18$, 95% CI $= [-0.05$ to $0.43]$. This suggested that between-challenge differences were more an artifact of variance in challenge brief length/specificity. Thus, to ensure meaningful comparability across challenges, we normalized the cosines by computing the z-score for each inspiration's cosine relative to other inspirations from the same challenge before analyzing the results in the full dataset. However, similar results are found using raw cosines, but with more uncertainty in the statistical coefficient estimates.

We then subtracted the cosine z-score from zero such that larger values meant more distant. From these 'reversed' cosine z-scores, two different distance

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

measures were computed to tease apart possibly distinct effects of source distance: 1) *max* distance ($DIST_{MAX}$), i.e., the distance of a concept's furthest source from the problem domain and 2) *mean* distance ($DIST_{MEAN}$) of the concept's sources. $DIST_{MAX}$ estimates 'upper bounds' for the benefits of distance: do the best ideas really come from the furthest sources? $DIST_{MEAN}$ capitalizes on the fact that many concepts relied on multiple inspirations and estimates the impact of the relative *balance* of relying on near vs. far sources (e.g., more near than far sources, or vice versa).

2.3.3 Control measures

Given our correlational approach, it is important to identify and rule out or adjust for other important factors that may influence the creativity of concepts (particularly in the later stages, where prototyping and feedback are especially important) and may be correlated with the predictor variables.

Feedback. Given the collaborative nature of OpenIDEO, we reasoned that feedback in the form of comments (labeled here as FEEDBACK) influences success. Comments can offer encouragement, raise issues/questions, or provide specific suggestions for improvement, all potentially significantly enhancing the quality of the concept. Further, feedback may be an alternate pathway to success via source distance, in that concepts that build on far sources may attract more attention and therefore higher levels of feedback, which then improve the quality of the concept.

Quality of cited sources. Concepts that build on existing high-quality concepts (e.g., those who end up being shortlisted or chosen as winners) have a particular advantage of being able to learn from the mistakes and shortcomings, good ideas, and feedback in these high-quality concepts. Thus, as a proxy measure of quality, the number of shortlisted concepts a given concept builds upon (labeled *SOURCESHORT*) could be a large determinant of a concept's success.

2.4 Analytic approach

We are interested in predicting the creative outcomes of 707 concepts, posted by 357 authors for 12 different design challenges. Authors are not cleanly nested within challenges, nor vice versa; our data are cross-classified, with concepts cross-classified within both authors and challenges (see Figure 6). This cross-classified structure violates assumptions of uniform independence between concepts: concepts posted by the same author or within the same challenge may be more similar to each other. Failing to account for this non-independence could lead to overestimates of the statistical significance of model estimates (i.e., make unwarranted claims of statistically significant effects). This issue is exacerbated when testing for small effects. Additionally, modeling between-author effects allows us to separate author-effects (e.g., higher/lower creativity) from the impact of sources on

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

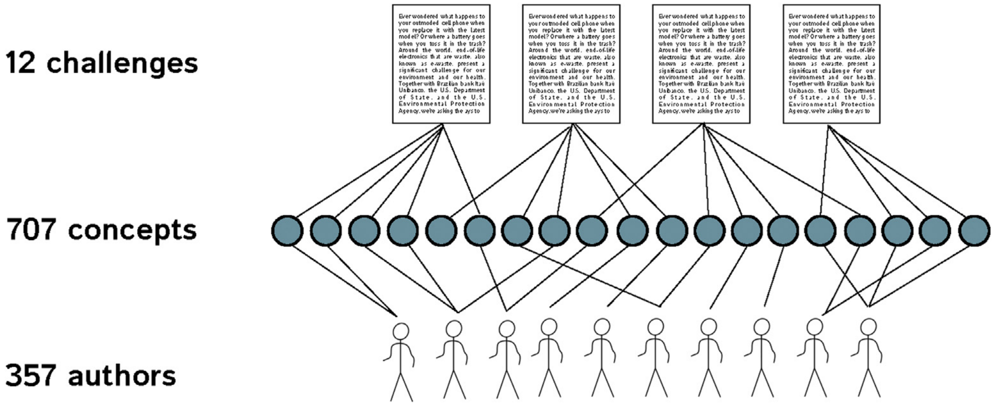

Figure 6 Illustrated cross-classified structure of the data

individual concepts Thus, we employ generalized linear mixed models (also called hierarchical generalized linear models) to model both fixed effects (of our independent and control variables) and random effects (potential variation of the outcome variable attributable to author- or challenge-nesting and also potential between-challenge variation in the effect of distance) on shortlist status (a binary variable, which requires logistic, rather than linear, regression).

An initial model predicting the outcome with only the intercept and between-challenge and -author variation confirms the presence of significant non-independence, with between-author and between-challenge variation in shortlist outcomes estimated at 0.44, and 0.50, respectively. The intra-class correlations for author-level and challenge-level variance in the intercept are ∼0.11 and 0.13, respectively, well above the cutoff recommended by Raudenbush and Bryk (2002).[1]

3 Results

3.1 Descriptive statistics

On average, 16% of concepts in the sample get shortlisted (see Table 2). $DIST_{MEAN}$ is centered approximately at 0, reflecting our normalization procedure. Both $DIST_{MAX}$ and $DIST_{MEAN}$ have a fair degree of negative skew. *SOURCESHORT* and *FEEDBACK* have strong positive skew (most concepts either have few comments or cite 0 or 1 shortlisted concepts).

There is a strong positive relationship between $DIST_{MAX}$ and $DIST_{MEAN}$ (see Table 3). All variables have significant bivariate correlations with *SHORTLIST* except for $DIST_{MAX}$; however, since it is a substantive variable of interest, we will model it nonetheless. Controlling for other variables might enable us to detect subtle effects.

Table 2 Descriptive statistics

Variable	*Valid* N	*Min*	*Max*	*Mean*	*Median*	*SD*
SHORTLIST	707	0.00	1.00	0.16	0.00	0.36
$DIST_{MAX}$	707	−3.85	1.90	0.45	0.76	0.85
$DIST_{MEAN}$	707	−3.85	1.67	−0.10	0.01	0.85
SOURCESHORT	707	0	11	0.51	0	0.96
FEEDBACK	707	0	67	8.43	6	9.45

3.2 Statistical models

We estimated separate models for the effects of $DIST_{MAX}$ and $DIST_{MEAN}$, each controlling for challenge- and author-nesting, *FEEDBACK*, and *SHORTSOURCE*.

3.2.1 Max distance

Our model estimated an inverse relationship between $DIST_{MAX}$ and Pr(shortlist), such that a 1-unit increase in $DIST_{MAX}$ predicted a 0.33 *decrease* in the log-odds of being shortlisted, after accounting for the effects of *FEEDBACK, SHORTSOURCE*, and challenge- and author-level nesting, $p < .05$ (see Appendix B for technical details on the statistical models). However, this coefficient was estimated with considerable uncertainty, as indicated by the large confidence intervals (coefficient could be as small as −0.06 or as large as −0.60); considering also the small bivariate correlation with *SHORTLIST*, we are fairly certain that the 'true' coefficient is *not* positive (*contra* the Conceptual Leap Hypothesis), but we are quite uncertain about its magnitude.

Figure 7 visually displays the estimated relationship between $DIST_{MAX}$ and Pr(shortlist), evaluated at mean values of feedback and shortlisted sources. To aid interpretation, we also plot the predicted Pr(shortlist) for concepts that cite no sources using a horizontal gray bar (bar width indicates uncertainty in estimate of Pr(shortlist)): concepts with approximately equivalent amounts of feedback (i.e., mean of 8.43), have a predicted Pr(shortlist = 0.09, 95% CI = [0.07−0.11]; using a logistic model, the coefficient for 'any citation' (controlling for feedback) is 0.31, 95% CI = [0.01−0.62]). This bar serves as an approximate 'control' group, allowing us to interpret the effect not just in terms of the effects of far sources relative to near sources, but also in comparison with using no sources. Comparing the fitted curve with this bar highlights how the

Table 3 Bivariate correlations

Variable	$DIST_{MAX}$	$DIST_{MEAN}$	SOURCESHORT	FEEDBACK
SHORTLIST	−0.05	−0.10*	0.11**	0.33***
$DIST_{MAX}$		0.77***	0.05	0.07^{m}
$DIST_{MEAN}$			−0.05	0.01
SOURCESHORT				0.12**

$^{m}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

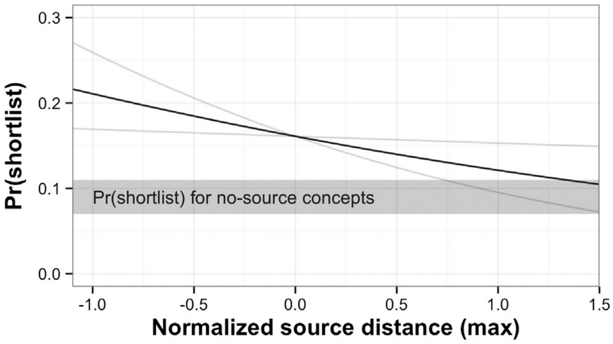

Figure 7 Model-fitted relationship between $\mathrm{DIST_{MAX}}$ *and Pr(shortlist), evaluated at mean values of feedback and source shortlist. Grayed lines are fits with upper and lower limits for 95% CI for effect of* $\mathrm{DIST_{MAX}}$

advantage of citing vs. not citing inspirations seems to be driven mostly by citing relatively near inspirations: Pr(shortlist) for concepts that cite far inspirations converges on that of no-citation concepts. We emphasize again that, despite the uncertainty in the *degree* of the negative relationship between $DIST_{MAX}$ and Pr(shortlist), the data do *not* support an inference that the best ideas are coming from the farthest inspirations: rather, relying on nearer rather than farther sources seems to lead to more creative design ideas. Importantly, this pattern of results was robust across challenges on the platform: the model estimated essentially zero between-challenge variation in the slope of $DIST_{MAX}$. $\chi^2(2) = 0.05$, $p = .49$ (see Figure 8).

3.2.2 Mean distance

Similar results were obtained for $DIST_{MEAN}$. There was a robust inverse relationship between $DIST_{MEAN}$ and Pr(shortlist), such that a 1-unit increase in $DIST_{MEAN}$ was associated with a *decrease* of approximately 0.40 in the log-odds of being shortlisted, $p < .05$. The estimates of this effect were obtained with similarly low precision regarding the magnitude of the effect, with 95% CI upper limit of at most $B = -0.09$ (but as high as -0.71). As shown in Figure 9, as $DIST_{MEAN}$ increases, Pr(shortlist) approaches that of non-citing concepts, again suggesting (as with $DIST_{MAX}$) that the most beneficial sources appear to be ones that are relatively close to the challenge domain. Again, as with $DIST_{MAX}$, this pattern of results did not vary across challenges: our model estimated essentially zero between-challenge variation in the slope of $DIST_{MEAN}$, $\chi^2(2) = 0.07$, $p = .48$ (see Figure 10).

4 Discussion

4.1 Summary and interpretation of findings

This study explored how the inspirational value of sources varies with their conceptual distance from the problem domain along the continuum from near to far. The study's findings provide no support for the notion that the best ideas come from building explicitly on the farthest sources. On the

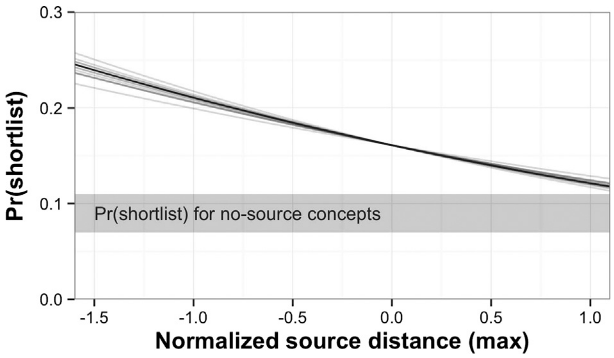

Figure 8 Overall and by-challenge model-fitted relationship between $\mathrm{DIST_{MAX}}$ *and Pr(shortlist). Fitted values evaluated at mean values of feedback and source shortlist. Grayed lines are fits for each individual challenge*

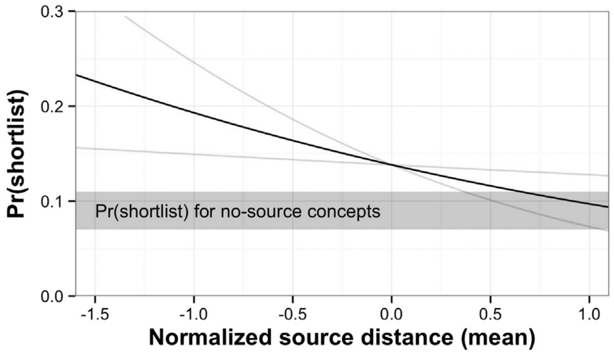

Figure 9 Model-fitted relationship between $\mathrm{DIST_{MEAN}}$ *and Pr(shortlist), evaluated at mean values of feedback and source shortlist. Grayed lines are fits with upper and lower limits for the 95% CI for the effect of* $\mathrm{DIST_{MEAN}}$

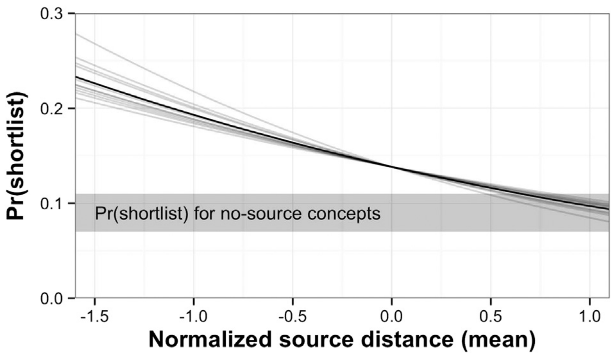

Figure 10 Overall and by-challenge model-fitted relationship between $\mathrm{DIST_{MEAN}}$ *and Pr(shortlist). Fitted values evaluated at mean values of feedback and source shortlist. Grayed lines are fits for each individual challenge*

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

Table 4 Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on $DIST_{MAX}$, with comparison to baseline model (controls only)

	Baseline model (controls only)	$DIST_{MAX}$, *fixed slope*	$DIST_{MAX}$, *random slope*
Fixed effects			
γ_{00}, intercept	−2.66 [−3.28, −2.03]	−2.57 [−3.29, −2.05]	−2.57 [−3.29, −2.05]
γ_{10}, *FEEDBACK*	0.09*** [0.07, 0.12]	0.10*** [0.07, 0.12]	0.10*** [0.07, 0.12]
γ_{20}, *SOURCESHORT*	0.14 [−0.08, 0.36]	0.15 [−0.07, 0.38]	0.15 [−0.07, 0.38]
γ_{30}, $DIST_{MAX}$		−0.33* [−0.60, −0.06]	−0.32* [−0.59, −0.06]
Random effects			
$u_{0authorj}$ for intercept	0.29	0.31	0.32
$u_{0challengek}$ for intercept	0.75	0.76	0.74
$u_{3challengek}$ for $DIST_{MAX}$			0.00
Model fit statistics			
Deviance	511.39	506.04	505.99
AIC	521.39	518.04	521.99

[m]$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$; 95% CI (Wald) = [lower, upper].

contrary, the benefits of building explicitly on inspirations seem to accrue mainly for concepts that build more on near than far inspirations. Importantly, these effects were consistently found in all of the challenges, addressing concerns raised about potential problem variation, at least among non-routine social innovation design problems.

4.2 Caveats and limitations

Some caveats should be discussed before addressing the implications of this study. First, the statistical patterns observed here are conditional: i.e., we find an inverse relationship between conceptual distance of *explicitly cited* inspiration sources and Pr(shortlist). Our data are silent on the effects of distance for concepts that did not cite sources (where lack of citation could indicate forgetting of sources or lack of conscious building on sources).

There is a potential concern over range restriction or attrition due to our reliance on self-identified sources. However, several features of the data help to ameliorate this concern. First, concepts that did not cite sources were overall of lower quality; thus, it is unlikely that the inverse effects of distance are solely due to attrition (e.g., beneficial far inspirations not being observed). Second, the integration of citations and building on sources into the overall OpenIDEO workflow and philosophy of ideation also helps ameliorate concerns about attrition of far sources. Finally, the dataset included many sources that were quite far away, providing sufficient data to statistically test the effects of relative reliance on far sources (even if they are overall under-reported). Nevertheless, we should still be cautious about making inferences about the impact of *unconscious* sources (since sources in this data are explicitly cited and therefore consciously built upon). However, as we note in the methods,

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

Table 5 Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on $DIST_{MEAN}$, with comparison to baseline model (controls only)

	Baseline model (controls only)	$\mathrm{DIST}_{\mathrm{MEAN}}$, *fixed slope*	$\mathrm{DIST}_{\mathrm{MEAN}}$, *random slope*
Fixed effects			
γ_{00}, intercept	−2.66 [−3.28, −2.03]	−2.74 [−3.36, −2.11]	−2.74 [−3.36, −2.11]
γ_{10}, *FEEDBACK*	0.09*** [0.07, 0.12]	0.10*** [0.07, 0.12]	0.10*** [0.07, 0.12]
γ_{20}, *SOURCESHORT*	0.14 [−0.08, 0.36]	0.13 [−0.09, 0.35]	0.13 [−0.09, 0.35]
γ_{30}, $DIST_{MEAN}$		−0.40* [−0.71, −0.09]	−0.40* [−0.73, −0.07]
Random effects			
$u_{0authorj}$ for intercept	0.29	0.31	0.30
$u_{0challengek}$ for intercept	0.75	0.73	0.73
$u_{3challengek}$ for $DIST_{MEAN}$			0.03
Model fit statistics			
Deviance	511.39	505.13	505.06
AIC	521.39	517.13	521.06

$^{m}p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; 95% CI (Wald) = [lower, upper].

the Conceptual Leap Hypothesis maps most cleanly to conscious inspiration processes (e.g., analogy).

Finally, some may be concerned that we have not measured novelty here. Conceivably, the benefits of distance may only be best observed for the novelty of ideas, and not necessarily quality, consistent with some recent work (Franke, Poetz, & Schreier, 2014). However, novelty *per se* does not produce creativity; we contend that to fully understand the effects of distance on design creativity, we must consider its impacts on both novelty and quality together (as our shortlist measure does).

4.3 Implications and future directions

Overall, our results consistently stand in opposition to the Conceptual Leap Hypothesis. In tandem with prior opposing findings (reviewed in the introduction), our work lends strength to alternative theories of inspiration by theorists like Perkins (1983), who argues that conceptual distance does not matter, and Weisberg (2009, 2011), who argues that within-domain expertise is a primary driver of creative cognition. We should be clear that our findings do not imply that *no* creative ideas come from far sources (indeed, in our data, some creative ideas did come from far sources); rather, our data suggest that the most creative design ideas are more likely to come from relying on a preponderance of nearer rather than farther sources. However, our data do suggest that highly creative ideas can often come from relying almost not at all on far sources (as evidenced by the analyses with maximum distance of sources). These good ideas may arise from iterative, deep search, a mechanism for creative breakthroughs that may be often overlooked but potentially at least as important as singular creative leaps (Chan & Schunn, 2014; Dow, Heddleston, &

Klemmer, 2009; Mecca & Mumford, 2013; Rietzschel, Nijstad, & Stroebe, 2007; Sawyer, 2012; Weisberg, 2011). In light of this and our findings, it may be fruitful to deemphasize the privileged role of far sources and mental leaps in theories of design inspiration and creative cognition.

How might this proposed theoretical revision be reconciled with the relatively robust finding that problem solvers from outside the problem domain can often produce the most creative ideas (Franke et al., 2014; Hargadon & Sutton, 1997; Jeppesen & Lakhani, 2010)? Returning to our reflections on the potential costs of processing far sources, one way to reconcile the two sets of findings might be to hypothesize that expertise in the distant source domain enables the impact of distant ideas by bypassing the cognitive costs of deeply understanding the far domain, and filters out shallow inferences that are not likely to lead to deep insights. Hargadon and Sutton's (1997) findings from their in-depth ethnographic study of the consistently innovative IDEO design firm are consistent with an expertise-mediation claim: the firm's cross-domain-inspired innovations appeared to flow at the day-to-day process level mainly from deep immersion of its designers in multiple disciplines, and 'division of expertise' within the firm, with brainstorms acting as crucial catalysts for involving experts from different domains on projects. However, studies directly testing expertise-mediation are scarce or non-existent.

Further, the weight of the present data, combined with prior studies showing no advantage of far sources, suggests that considering alternative mechanisms of outside-domain advantage may be more theoretically fruitful: for instance, perhaps the advantage of outside-domain problem-solvers arises from the different perspectives they bring to the problem — allowing for more flexible and alternative problem representations, which may lead to breakthrough insights (Kaplan & Simon, 1990; Knoblich, Ohlsson, Haider, & Rhenius, 1999; Öllinger, Jones, Faber, & Knoblich, 2012). Domain-outsiders may also have a looser attachment to the *status quo* or prior successful solutions by virtue of being a 'newcomer' to the domain (Choi & Levine, 2004) — leading to higher readiness to consider good ideas that challenge existing assumptions within the domain — rather than knowledge and transfer of different solutions *per se*.

Finally, it would be interesting to examine potential moderating influences of source processing *strategies*. In our data, closer sources were more beneficial, but good ideas also did come from far sources; however, as we have argued, it can be more difficult to convert far sources into viable concepts. Are there common strategies for effective conversion of far sources, and are they *different* from strategies for effectively building on near sources? For example, one effective strategy for building on sources while avoiding fixation is to use a schema-based strategy (i.e., extract and transfer abstract functional principles rather than concrete solution features; Ahmed & Christensen, 2009; Yu, Kraut, & Kittur, 2014). Are there processing strategies that expert creative

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

designers apply uniquely to far sources (e.g., to deal with potentially unalignable differences)? Answering this question can shed further light on the variety of ways designers can be inspired by sources to produce creative design ideas.

We close by noting the methodological contribution of this work. While we are not the first to use topic modeling to explore semantic meaning in a large collection of documents, we are the first to our knowledge to validate this method in the context of large-scale study of design ideas. We have shown that the topic model approach adequately captures human intuitions about the semantics of the design space, while providing dramatic savings in cost: indeed, such an approach can make more complex research questions (e.g., exploring pairwise distances between design idea or, tracing conceptual paths/moves in a design ideation session) much more feasible without sacrificing too much quality. We believe this approach can be a potentially valuable way for creativity researchers to study the dynamics of idea generation at scale, while avoiding the (previously inevitable) tradeoff between internal validity (e.g., having adequate statistical power) and external validity (e.g., using real, complex design problems and ideas instead of toy problems).

Appendix A. Topic model technical details

A.1. Document preprocessing

All documents were first tokenized using the TreeBank Tokenizer from the open-source Natural Language Tool Kit Python library (Bird, Klein, & Loper, 2009). To improve the information content of the document text, we removed a standard list of stopwords, i.e., highly frequent words that do not carry semantic meaning on their own (e.g., 'the', 'this'). We used the open-source MAchine Learning for LanguagE Toolkit's (MALLET; McCallum, 2002) stopword list.

A.2. Model parameter selection

We used MALLET to train our LDA model, with asymmetric priors for the topic-document and topic-word distributions, which allows for some words to be more prominent than others and some topics to be more prominent than others, typically improving model fit and performance (Wallach, Mimno, & McCallum, 2009). Priors were optimized using MALLET's in-package optimization option.

LDA requires that K (the number of topics) be prespecified by the modeler. Model fit typically improves with K, with diminishing returns past a certain point. Intuitively, higher K leads to finer-grained topical distinctions, but too high K may lead to uninterpretable topics; on the other hand, too low K would yield too general topics. Further, traditional methods of optimizing K (computing 'perplexity', or the likelihood of observing the distribution of

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

words in the corpus given a topic model of the corpus) do not always correlate with human judgments of model quality (e.g., domain expert evaluations of topic quality; Chang, Gerrish, Wang, Boyd-graber, & Blei, 2009).

We explored the following settings of K: [12, 25, 50, 100, 200, 300, 400, 500, 600, 700]. Because the optimization algorithm for the prior parameters is nondeterministic, models with identical K might produce noticeably different topic model solutions, e.g., if the optimization search space is rugged, the algorithm might get trapped in different local maxima. Therefore, we ran 50 models at each K, using identical settings (i.e., 1000 iterations of the Gibbs sampler, internally optimizing parameters for the asymmetric priors). Figure 11 shows the mean fit (with both continuous and binary similarity judgments) at each level of K.

Model fit is generally fairly high at all levels of K, with the continuous judgments tending to increase very slightly with K, tapering out past 400. Fit with binary judgments tended to decrease (also very slightly) with K, probably reflecting the decreasing utility of increasingly finer-grained distinctions for a binary same/different classification. Because we wanted to optimize for fit with human judgments of conceptual distance overall, we selected the level of K at which the divergent lines for fit with continuous and binary judgments first begin to cross (i.e., at $K = 400$). Subsequently, we created a combined 'fit' measure (sum of the correlation coefficients for fit vs. continuous and binary judgments), and selected the model with $K = 400$ that had the best overall fit measure. However, as we report in the next section, the results of our analyses are robust to different settings of K.

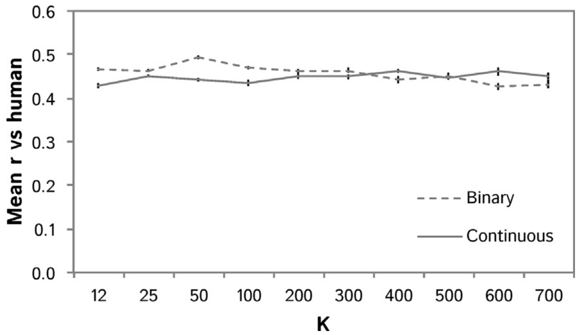

Figure 11 Mean fit (with ±1 SE) vs. human judgments for LDA cosines by level of K

Appendix B. Statistical modeling technical details

B.1. Statistical modeling approach

All models were fitted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2013) in R (R Core Team, 2013), using full maximum likelihood estimation by the Laplace approximation. The following is the general structure of these models (in mixed model notation):

$$\eta_{i(authorjchallengek)} = \gamma_{00} + \sum_{q} \gamma_{q0} X_{qi} + u_{0authorj} + u_{0challengek}$$

where

- $\eta_{i(authorjchallengek)}$ is the predicted log odds of being shortlisted for the ith concept posted by the *j*th author in the *k*th challenge
- γ_{00} is the grand mean log odds for all concepts
- γ_{q0} is a vector of q predictors ($q = 0$ for our null model)
- $u_{0authorj}$ and $u_{0challengek}$ are the random effects contribution of variation between-authors and between-challenges for mean γ_{00} (i.e., how much a given author or challenge varies from the mean)

A baseline model with only control variables and variance components was first fitted. Then, for the models for both $DIST_{MAX}$ and $DIST_{MEAN}$, we first estimated a model with a fixed effect of distance, and then a random effect (to test for problem variation). These random slopes models include the additional parameter $u_{1challengek}$ that models the between-challenge variance component for the slope of distance.

B.2. Model selection

Estimates and test statistics for each step in our model-building procedure are shown in Tables 4 and 5. We first fitted a model predicting Pr(shortlist) with our control variables to serve as a baseline for evaluating the predictive power of our distance measures. The baseline model estimates a strong positive effect of *FEEDBACK*, estimated with high precision: each additional comment added 0.10 [0.07, 0.12] to the log-odds of being shortlisted, $p < .001$. The model also estimated a positive effect of *SHORTSOURCE*, $B = 0.14$ [−0.08, 0.36] but with poor precision, and falling short of conventional statistical significance, $p = .21$; nevertheless, we leave it in the model for theoretical reasons. The baseline model is a good fit to the data, reducing deviance from the null model (with no control variables) by a large and statistically significant amount, $\chi^2(1) = 74.35$, $p = .00$.

For the fixed slope model for $DIST_{MAZ}$, adding the coefficient for results in a significant reduction in deviance from the baseline model, $\chi^2(2) = 0.13$, $p = .47$. The random slope model did not significantly reduce deviance in comparison with the simpler fixed slope model, $\chi^2(2) = 0.05$, $p = .49$ (p-value is halved, heeding common warnings that a likelihood ratio test discriminating two models that differ on only one variance component may be overly conservative, e.g., Pinheiro & Bates, 2000). Also, the Akaike Information Criterion (AIC) increases from the fixed to random slope model. Thus, we select the fixed slope model (i.e., no problem-variation) as our best estimate of the effects of $DIST_{MAX}$. This final model has an overall deviance reduction vs. null at $\chi^2(3) = 79.71$, $p = .00$.

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

We used the same procedure for model selection for the $DIST_{MEAN}$ models. The fixed slope model results in a small but significant reduction in deviance from the baseline model, $\chi^2(1) = 6.27$, $p = .01$. Adding the variance component for the slope of $DIST_{MEAN}$ increases the AIC, and does not significantly reduce deviance, $\chi^2(2) = 0.07$, $p = .48$ (again, p-value here is halved to correct for overconservativeness). Thus, again we select the fixed slope model as our final model for the effects of $DIST_{MEAN}$. This final model has an overall reduction in deviance from the null model of about $\chi^2(3) = 80.61$, $p = .00$.

B.3. Robustness and sensitivity

We tested the robustness of our coefficient estimates by calculating outlier influence statistics using the influence.measures method in the stats package in R, applied to logistic regression model variants of both the $DIST_{MEAN}$ and $DIST_{MAX}$ models (i.e., without author- and challenge-level variance components; coefficient estimates are almost identical to the fixed slope multilevel models): DFBETAS and Cook's Distance measures were below recommended thresholds for all data points (Fox, 2002).

Addressing potential concerns about sensitivity to topic model parameter settings, we also fitted the same fixed slope multilevel models using recomputed conceptual distance measures for the top 20 (best-fitting) topic models at $K = 200$, 300, 400, 500, and 600 (total of 100 models). All models produced negative estimates for the effect of both $DIST_{MEAN}$ and $DIST_{MAX}$, with poorer precision for lower K. Thus, our results are robust to different settings of K for the topic models.

We also address potential concerns about interactions with expertise by fitting a model that allowed the slope of distance to vary by authors. In this model, the overall mean effect of distance remained almost identical ($B = -0.46$), and the model's fit was not significantly better than the fixed slope model, $\chi^2(3) = 3.44$, $p = .16$, indicating a lack of statistically significant between-author variability for the slope of distance.

Finally, we also fitted models that considered not just immediately cited inspirations, but also indirectly cited inspirations (i.e., inspirations cited by cited inspirations), and they too yielded almost identical coefficient estimates and confidence intervals.

References

Ahmed, S., & Christensen, B. T. (2009). An in situ study of analogical reasoning in novice and experienced designer engineers. *Journal of Mechanical Design, 131*(11), 111004.

Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997−1013.

Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In C. S. Schreiner (Ed.), *Handbook of research on assessment* *technologies, methods, and applications in higher education* (pp. 65–77), Hershey, PA.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *Lme4: Linear mixed-effects models using eigen and S4. R package version 1.0-5*. [Computer software]. Retrieved from. http://CRAN.R-project.org/package=lme4.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python*. O'Reilly Media Inc.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on machine learning* (pp. 113–120).

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics* 17–35.

Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research* 993–1022.

Brown, R. T. (1989). Creativity: what are we to measure? In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 3–32), New York, NY.

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(3), 432–442.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. http://dx.doi.org/10.1038/nrn3475.

Chakrabarti, A. (2006). Defining and supporting design creativity. In *Proceedings of the 9th international design conference DESIGN 2006* (pp. 479–486).

Chan, J., Fu, K., Schunn, C. D., Cagan, J., Wood, K. L., & Kotovsky, K. (2011). On the benefits and pitfalls of analogies for innovative design: ideation performance based on analogical distance, commonness, and modality of examples. *Journal of Mechanical Design, 133*, 081004.

Chang, J., Gerrish, S., Wang, C., Boyd-graber, J. L., & Blei, D. M. (2009). Reading tea leaves: how humans interpret topic models. *Advances in neural information processing systems* 288–296.

Chan, J., & Schunn, C. (2014). The impact of analogies on creative concept generation: lessons from an in vivo study in engineering design. *Cognitive Science*. http://dx.doi.org/10.1111/cogs.12127.

Chiu, I., & Shu, H. (2012). Investigating effects of oppositely related semantic stimuli on design concept creativity. *Journal of Engineering Design, 23*(4), 271–296. http://dx.doi.org/10.1080/09544828.2011.603298.

Choi, H. S., & Levine, J. M. (2004). Minority influence in work teams: the impact of newcomers. *Journal of Experimental Social Psychology, 40*(2), 273–280.

Dahl, D. W., & Moreau, P. (2002). The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research, 39*(1), 47–60.

Deerwester, S., Dumais, S. T., Furnas, G. W., & Landauer, T. K. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 1990.

Dow, S. P., Heddleston, K., & Klemmer, S. R. (2009). The efficacy of prototyping under time constraints. In *Proceedings of the 7th ACM conference on creativity and cognition*.

Dunbar, K. N. (1997). How scientists think: on-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative* *thought: An investigation of conceptual structures and processes* (pp. 461–493), Washington, D.C.

Dym, C. L. (1994). *Engineering design: A synthesis of views*. New York, NY: Cambridge University Press.

Eckert, C., & Stacey, M. (1998). Fortune favours only the prepared mind: why sources of inspiration are essential for continuing creativity. *Creativity and Innovation Management, 7*(1), 1–12.

Enkel, E., & Gassmann, O. (2010). Creative imitation: exploring the case of cross-industry innovation. *R & D Management, 40*(3), 256–270.

Findlay, A. (1965). *A hundred years of chemistry* (3rd ed.). London: Duckworth.

Fox, J. (2002). *An R and s-plus companion to applied regression*. Sage.

Franke, N., Poetz, M. K., & Schreier, M. (2014). Integrating problem solvers from analogous markets in new product ideation. *Management Science, 60*(4), 1063–1081.

Freeman, A., & Golden, B. (1997). *Why didn't I think of that? Bizarre origins of ingenious inventions we couldn't live without*. New York: John Wiley.

Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., & 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), 021007. http://dx.doi.org/10.1115/1.4023158.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*(1), 45–56.

German, T. P., & Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. *Psychological Science, 16*(1), 1–5.

Gero, J. S. (2000). Computational models of innovative and creative design processes. *Technological Forecasting and Social Change, 64*(2), 183–196.

Goldschmidt, G., & Smolkov, M. (2006). Variances in the impact of visual stimuli on design problem solving performance. *Design Studies, 27*(5), 549–569.

Gonçalves, M., Cardoso, C., & Badke-Schaub, P. (2013). Inspiration peak: exploring the semantic distance between design problem and textual inspirational stimuli. *International Journal of Design Creativity and Innovation*1–18, (ahead-of-print).

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl. 1), 5228–5235. http://dx.doi.org/10.1073/pnas.0307752101.

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly, 42*(4), 716. http://dx.doi.org/10.2307/2393655.

Helms, M., Vattam, S. S., & Goel, A. K. (2009). Biologically inspired design: process and products. *Design Studies, 30*(5), 606–622.

Hender, J. M., Dean, D. L., Rodgers, T. L., & Jay, F. F. (2002). An examination of the impact of stimuli type and GSS structure on creativity: brainstorming versus non-brainstorming techniques in a GSS environment. *Journal of Management Information Systems, 18*(4), 59–85.

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps: Analogy in creative thought*. Cambridge, MA.

Impact Stories. (n.d.). Impact stories. [Web page]. Retrieved from http://www.openideo.com/content/impact-stories.

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies, 12*(1), 3–11.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization Science, 21*(5), 1016–1033.

Jessup, E. R., & Martin, J. H. (2001). Taking a new look at the latent semantic analysis approach to information retrieval. *Computational information retrieval*. Philadelphia: SIAM 121−144.

Kalogerakis, K., Lu, C., & Herstatt, C. (2010). Developing innovations based on analogies: experience from design and engineering consultants. *Journal of Product Innovation Management, 27*, 418−436.

Kaplan, C., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374−419.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1534−1555.

Liikkanen, L. A., & Perttula, M. (2008). Inspiring design idea generation: insights from a memory-search perspective. *Journal of Engineering Design, 21*(5), 545−560.

Linsey, J., Tseng, I., Fu, K., Cagan, J., Wood, K., & Schunn, C. (2010). A study of design fixation, its mitigation and perception in engineering design faculty. *Journal of Mechanical Design, 132*(4). 041003-1-12.

Malaga, R. A. (2000). The effect of stimulus modes and associative distance in individual creativity support systems. *Decision Support Systems, 29*(2), 125−141.

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 886−897.

Marsh, R. L., Ward, T. B., & Landau, J. D. (1999). The inadvertent use of prior knowledge in a generative cognitive task. *Memory & Cognition, 27*(1), 94−105.

McCallum, A. K. (2002). MALLET: a machine learning for language toolkit. [Computer software]. Retrieved from. http://mallet.cs.umass.edu.

Mecca, J. T., & Mumford, M. D. (2013). Imitation and creativity: beneficial effects of propulsion strategies and specificity. *The Journal of Creative Behavior*. http://dx.doi.org/10.1002/jocb.49.

Öllinger, M., Jones, G., Faber, A. H., & Knoblich, G. (2012). Cognitive mechanisms of insight: the role of heuristics and representational change in solving the eight-coin problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. http://dx.doi.org/10.1037/a0029194.

Perkins, D. N. (1983). Novel remote analogies seldom contribute to discovery. *The Journal of Creative Behavior, 17*(4), 223−239.

Perkins, D. N. (1997). Creativity's camel: the role of analogy in invention. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 523−538). Washington, D.C.: American Psychological Association.

Pinheiro, J. C., & Bates, D. M. (2000). *Linear mixed-effects models: Basic concepts and examples*. Springer.

Poze, T. (1983). Analogical connections — the essence of creativity. *The Journal of Creative Behavior, 17*(4), 240−258.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA.

R Core Team. (2013). *R: A language and environment for statistical computing*. [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from. http://www.R-project.org/.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: the effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology, 43*(6), 933−946.

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

Sawyer, R. K. (2012). *Explaining creativity: The science of human innovation* (2nd ed.). New York: Oxford University Press.
Schunn, C. D., & Dunbar, K. N. (1996). Priming, analogy, and awareness in complex reasoning. *Memory & Cognition, 24*(3), 271–284.
Schwartz, A. H., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., et al. (2013). Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS ONE, 8*(9), e73791.
Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 424–440). New York, NY: Lawrence Erlbaum.
Tseng, I., Moss, J., Cagan, J., & Kotovsky, K. (2008). The role of timing and analogical similarity in the stimulation of idea generation in design. *Design Studies, 29*(3), 203–221.
Ullman, D. (2002). *The mechanical design process*. New York, NY (3rd ed.).
Wallach, H. M., Mimno, D. M., & McCallum, A. (2009). Rethinking LDA: why priors matter. In*NIPS, Vol 22* (pp. 1973–1981).
Ward, T. B. (1994). Structured imagination: the role of category structure in exemplar generation. *Cognitive Psychology, 27*(1), 1–40.
Ward, T. B. (1998). Analogical distance and purpose in creative thought: mental leaps versus mental hops. In K. J. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 221–230), Sofia, Bulgaria.
Weisberg, R. W. (2009). On "out-of-the-box" thinking in creativity. In A. B. Markman, & K. L. Wood (Eds.), *Tools for innovation* (pp. 23–47), New York, NY.
Weisberg, R. W. (2011). Frank lloyd wright's fallingwater: a case study in inside-the-box creativity. *Creativity Research Journal, 23*(4), 296–312. http://dx.doi.org/10.1080/10400419.2011.621814.
Wiley, J. (1998). Expertise as mental set: the effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*(4), 716–730.
Wilson, J. O., Rosen, D., Nelson, B. A., & Yen, J. (2010). The effects of biological examples in idea generation. *Design Studies, 31*(2), 169–186.
Yu, L., Kraut, B., & Kittur, A. (2014). Distributed analogical idea generation: innovating with crowds. In *Proceedings of the ACM conference on human factors in computing systems (CHI'14)*.
Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1049–1060.

Endnote

Although concept-level variance is not estimated in mixed logistic regressions, we follow Zeger, Liang, and Albert's (1988) suggestion of $(15/16)\pi^3/3$ as a reasonable approximation for residual level-1 variance (the concept level in our case).

Please cite this article in press as: Chan, J., et al., Do the best design ideas (really) come from conceptually distant sources of inspiration?, Design Studies (2014), http://dx.doi.org/10.1016/j.destud.2014.08.001

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2014) 1–30

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12127

The Impact of Analogies on Creative Concept Generation: Lessons From an *In Vivo* Study in Engineering Design

Joel Chan, Christian Schunn

Learning Research and Development Center, University of Pittsburgh

Received 14 September 2012; received in revised form 17 September 2013; accepted 26 September 2013

Abstract

Research on innovation often highlights analogies from sources outside the current problem domain as a major source of novel concepts; however, the mechanisms underlying this relationship are not well understood. We analyzed the temporal interplay between far analogy use and creative concept generation in a professional design team's brainstorming conversations, investigating the hypothesis that far analogies lead directly to very novel concepts via large steps in conceptual spaces (jumps). Surprisingly, we found that concepts were more similar to their preceding concepts after far analogy use compared to baseline situations (i.e., without far analogy use). Yet far analogies increased the team's concept generation rate compared to baseline conditions. Overall, these results challenge the view that far analogies primarily lead to novel concepts via jumps in conceptual spaces and suggest alternative pathways from far analogies to novel concepts (e.g., iterative, deep exploration within a functional space).

Keywords: Analogy; Creativity; Design cognition; Problem solving; *In vivo*

1. Introduction

Innovation is a key output of human cognition and therefore an important object of study for cognitive science. Arguably, the ability to produce novel artifacts that solve some problem and bring significant value to stakeholders/society is comparable to other great human intellectual achievements, such as great art, literature, and achieving detailed understanding of the natural world through the scientific method. Consider the LIFE-SAVER® portable filtration system, a durable, inexpensive, and portable means of turning dirty and pathogen-ridden water into clean, life-saving drinkable water in seconds

Correspondence should be sent to Joel Chan, Learning Research and Development Center, University of Pittsburgh, LRDC 823, 3939 O'Hara Street, Pittsburgh, PA 15260. E-mail: joc59@pitt.edu

(Walters, 2013). It represents a viable solution to the extensive problem of water poverty, sidestepping the major obstacle of infrastructure modification difficulties; in fact, it is already transforming the lives of thousands of people in rural Borneo. How do innovations like this arise from human minds and their interactions with their surroundings?

From both a practical and a theoretical standpoint, the mental representations and processes that lead to innovation are a worthy topic of inquiry for cognitive science; practically, because of innovation's cultural and economic importance, and theoretically, because by virtue of its complex, multifaceted nature, it can serve as a test bed for theories of cognition. In decades of cognitive-based research on the topic of innovation, researchers and theorists have uncovered the importance of collaboration and serendipity (Sawyer, 2007), incubation (Christensen & Schunn, 2005; Seifert, Meyer, Davidson, Patalano, & Yaniv, 1995; Tseng, Moss, Cagan, & Kotovsky, 2008), external representations (Goel, 1995), and mental simulation (Ball & Christensen, 2009; Christensen & Schunn, 2009b), among others. Fundamental to innovation, however, is concept generation. One cannot "make a silk purse out of a sow's ear" (Kornish & Ulrich, 2014); execution and implementation are critical, but innovation ultimately begins with good concepts. More specifically, as some theorists would argue, "breakthrough" or "radical" innovation comes from good concepts that are also very *new* (Boden, 2004).

The present work focuses on analogy, a cognitive process that has been hypothesized to be a major source of new concepts. Analogy is a fundamental cognitive process in which a *source* and *target* domain of knowledge are linked to one another by a systematic mapping of attributes and relations, which then allows for transfer of knowledge to the target (French, 2002; Gentner, 1983; Gentner & Forbus, 2011; Holyoak & Thagard, 1996; Hummel & Holyoak, 1997). Theoretical accounts of analogy describe it as a central cognitive mechanism for bridging seemingly disparate conceptual spaces, enabling thinking across categories and implicit conceptual boundaries (Gentner, 2003; Hofstadter, 2001; Holyoak & Thagard, 1996). This process appears to be important for generating novel concepts in a wide variety of domains, perhaps most prominently in scientific discovery (Clement, 1988; Dunbar, 1997; Gentner et al., 1997; Holyoak & Thagard, 1996; Nersessian, 1992; Oppenheimer, 1956) and—the domain of focus in this article—technological invention and innovation. In technological innovation, analogies have been associated with innovative outcomes in protocol studies and retrospective studies of expert and prominent inventors and designers (Carlson & Gorman, 1990; Gorman, 1997), experimental studies of design processes (Chan et al., 2011; Dahl & Moreau, 2002; Goldschmidt, 2001; Vargas-Hernandez, Shah, & Smith, 2010), and computational models of design (Gero & Kazakov, 1998). Analogy is also an important component of formal innovative design methods, such as design-by-analogy (French, 1988; Gordon, 1961; Hacco & Shu, 2002; Hey, Linsey, Agogino, & Wood, 2008; Linsey, Murphy, Laux, Markman, & Wood, 2009).

Not all analogies are thought to be equally productive for creative outcomes. Many theorists argue that, when considering the analogical distance of sources, far analogies—that is, from sources that have a low degree of overlap of surface elements with the current problem domain—hold the most potential for generating very new concepts (Gentner

& Markman, 1997; Holyoak & Thagard, 1996; Poze, 1983; Ward, 1998). A number of studies have shown that using or being stimulated by far analogies can increase production of very new concepts relative to near or no analogies (Chan et al., 2011; Chiu & Shu, 2012; Dahl & Moreau, 2002; Gonçalves, Cardoso, & Badke-Schaub, 2013; Hender, Dean, Rodgers, & Jay, 2002), although some studies have not replicated this finding (Huh & Kim, 2012; Malaga, 2000; Wilson, Rosen, Nelson, & Yen, 2010).

How might far analogies lead to very novel concepts? One prominent hypothesis, borrowing from the theoretical characterization of creative concept generation as search in a space (Boden, 2004; Goel & Pirolli, 1992; Perkins, 1994, 1997; Simon, 1996), is that far analogies enable "jumps" in the space of possible concepts. In other words, in contrast to more incremental search strategies, such as hill climbing, using far analogies enables the creator to "jump" to concepts that are very different from the set of concepts currently considered. The early roots of this notion can be found in Koestler's (1964) "bisociation" theory of creativity, where he argues that the best concepts come from when two previously unrelated concepts are combined into a new concept that is highly original and different form current concepts. Mednick's (1962) associative theory of creativity advances a similar argument about far connections enabling jumps in associative space to a highly creative concept. More recently, Perkins (1994, 1997) outlined the "canyon" problem as a topographical challenge of search spaces for problems requiring innovation, where the crucial insight may lie in a very distant part of the space, isolated from one's current location; importantly, he suggests that "analogy inherently has the power to step across canyons by relating one domain to another" (Perkins, 1997, p. 534). The idea that crucial insights may lie outside one's domain is consistent with the rise of collaborations and interdisciplinarity in science and technology (Jones, 2009; Paletz & Schunn, 2010; Wuchty, Jones, & Uzzi, 2007). Social network theories of innovation also emphasize the privileged position of agents positioned in "structural holes" in the information network (Burt, 2004; Hargadon, 2002; Ruef, 2002; Tortoriello & Krackhardt, 2010), being able to bridge knowledge and resources from structurally separated regions of the network.

While this hypothesis about the relationship between far analogies and creative concept generation (i.e., far analogies lead to very novel concepts via "jumps" in conceptual space) seems plausible and theoretically motivated, there is a critical empirical gap; online studies of concept generation have not measured and analyzed far analogy use and conceptual search patterns together. Prior studies showing a positive effect of far analogies on novelty of generated concepts have typically done so in an "input-output" design, where the conceptual *outputs* of designers who are given analogies as stimulation are compared to those of designers who are not given analogies. The lack of "online" process data still leaves open the possibility that the designers in the analogy groups may be chaining together far analogies and generated concepts to incrementally arrive at novel concepts in a way that is not recorded in their final recorded designs. Retrospective interviews of prominent innovators are of little help; potential issues surrounding incompleteness, inaccuracy, and bias in retrospective reports are well documented (for a review, see Schacter, 1999) and may be exacerbated when one is asked to retrospect for a phenomenon about which one has (lay) theories, as may often be the case in creativity research (Perkins, 1981). This lack of detailed examinations of the interplay between far analogies and concept generation is a major obstacle to theoretical progress in understanding the precise ways in which analogies can impact creative concept generation. Taking an *in vivo* approach, we address this gap by presenting detailed analyses of the online interplay between far analogies and concept generation in a team of real-world professional designers.

2. Study 1

2.1. Overview

This study presents analyses of multiple hours of naturalistic brainstorming conversations of a real-world professional design team, involving a large number of analogies and diverse set of subproblems. The design team consisted of 10 professionals from a range of design-related disciplines, including electronics and business development, mechanical engineering, business consulting, ergonomics and usability, and industrial design and project management. The team was tasked with developing a new product concept for a hand-held application of thermal printing technology for children. Within a larger taxonomy of design problems, ranging from routine (e.g., configuration/parametric design) to non-routine (creating original products), where non-routine problems are perceived as requiring more innovation (Chakrabarti, 2006; Dym, 1994; Gero, 2000), this problem is clearly non-routine, with the goal being to design a completely novel product in a new market, albeit leveraging an existing core technology. Thus, this design context is well suited for observation of processes that might lead to more "radical" rather than "incremental" innovation, where radical innovation has been more closely identified as coming from very novel concepts (Dewar & Dutton, 1986).

These conversations unfolded over the course of two design team meetings, structured as "brainstorms," with a focus on concept generation; the first meeting lasted 1 h and 37 min and focused on mechanical design subproblems; the second meeting lasted 1 h and 40 min and focused on electronics subproblems. The meetings were recorded with four pre-placed cameras in the meeting room. Although no researcher was present at either meeting, the designers were aware that they were being recorded, and that the data would be used, along with recordings of design meetings at other companies, for a large study by the Open University on "design meetings in practice." The transcripts include humor and outlandish statements, suggesting they were not very inhibited by the presence of cameras.

Prior to the first meeting, the designers received a design brief that requested that they think about problems related to the print head mounting design and pen format (e.g., keeping the print head level in spite of users' wobbly arm movement, protecting the print head from overheating and impact damage). To stimulate concept generation for these problems, the designers were also asked to bring along products (or pictures of products) that glide smoothly over contours.

The purpose of Study 1 was to determine whether far analogies were associated with conceptual jumps during concept generation. We operationalized jumps in terms of functional distance, that is, the degree to which a given concept's described functionality (i.e., a way of satisfying some design requirement vs. changes in color or manufacturing material not directly tied to changed functionality) was different from a prior concept or set of concepts. This operationalization reflects our focus on "radical" innovation, which in engineering and technological contexts has been associated with changes in functionality; for instance, Sood and Tellis (2005) argue that "platform innovation"—new functionality based on novel working/scientific principles (e.g., from magnetism for reading/writing data with floppy disks to laser optics for compact disks)—is where "breakthrough" or "radical" innovation happens. With this operationalization, the working hypothesis to be tested in Study 1 was the following: *The functional distance of a proposed concept from concepts recently considered will be reliably greater when preceded by far analogies versus baseline, that is, when not preceded by far analogies.*

Some discussion of validity and reliability is required given the deviations from a typical laboratory study along several dimensions. First, the data are narrow in the sense of studying one team and only 10 individuals working on one larger design problem. But the team worked on many different functional problems and generated a large number of different analogies; thus, this dataset is broader in another sense than a typical laboratory study that often examines the effect of one or two provided analogies on one given problem. Second, in terms of generalizability, it is not obvious that studying 100 undergraduates with low prior knowledge in the given domain, little relevant disciplinary training, and little incentive to do well produces outcomes of greater generalizability than the study of seven motivated, knowledgeable, and richly trained adults from diverse backgrounds working over multiple hours on many subproblems. Instead, it is likely that cognitive science will benefit from encouraging just as many studies of cognition in the wild as studies in the laboratory.

2.2. Methods

2.2.1. Segmentation

Analysis was conducted on the transcribed audio from the two meetings. Transcripts were segmented into lines by utterances, such that each line contained a separate thought; in this segmentation, a single sentence or speaker turn could span multiple lines. The segmentation procedure resulted in a total of 4,594 lines, 2,382 in the first meeting and 2,212 in the second.

2.2.2. Coding analogy use

Coding of analogy use was conducted by a prior research team, whose findings have been published in Ball and Christensen (2009); the second author, who has many years of expertise in studying analogy *in vivo*, served as the primary coder, with a secondary coder not affiliated with the research project recruited and trained to serve as a reliability check. Analogies were coded at the sentence/turn level but tagged at the line level, meaning that analogies often spanned multiple lines. Sentences were coded as analogies any time a designer referred to another source of knowledge and attempted to transfer concepts from that source to the target domain. One Hundred and forty-four analogies were found across the two transcripts (79 in the first and 65 in the second), with all designers contributing analogies at approximately the same rate, commensurate with their level of participation in the meetings overall (correlation between number of analogy and non-analogy utterances across designers was high, $r = .72$). Inter-rater reliability, assessed by comparing the primary and secondary coder's codes for approximately 1 h worth of transcript, was acceptable, at (Cohen's kappa) $k = .77$. This method of assessing inter-rater reliability was also used for the remaining analogy codes.

Analogies were coded for both distance and purpose. Following previous *in vivo* studies of analogy (Ball & Christensen, 2009; Christensen & Schunn, 2007), analogies were coded near versus far as follows: *Near* analogies involved mappings from sources that related to tools, mechanisms, and processes associated with graphical production and printing, while *far* analogies involved mappings from more far sources (see Tables 1 and 2). Of the 144 analogies found, 16% were coded as near, and 84% were coded as far. Inter-rater reliability was very high, $k = .99$. Because near analogies were relatively rare and because they are not the focus of prior hypotheses regarding impacts on concept generation, the analyses focus on the effects of the far analogies.

Following previous work (Ball & Christensen, 2009; Blanchette & Dunbar, 2001; Christensen & Schunn, 2007), analogical purpose (i.e., the goal or function of the analogy) was coded at three levels, with a fourth level added as a theoretical contribution by Ball and Christensen (2009; see Tables 3–6 for examples): (a) *Problem identification*—noticing a possible problem in the emerging design, where the problem was taken from an analogous source domain; (b) *Concept generation*—transferring possible design concepts from the source domain to the target domain; (c) *Explanation*—using a concept from the source domain to explain some aspect of the target domain to members of the design team; and (d) *Function-finding*—active mapping of new functions to the design form currently being developed (i.e., a thermal printing pen). Inter-rater reliability for this coding scheme was also high, $k = .85$.

2.2.3. Coding concept generation

Because coding concept generation was more difficult, three coders, including the author and two trained research assistants, identified generated concepts and the subproblems they were intended to address. Similar to the coding of analogy use, concepts were

Table 1
Example of near analogy

976	Alan	the other thing to think about is
977		in almost all cases when I look at pens the apart from re-wired sort of micropens the th- tip is actually the narrowest part of the product
988		whereas in what we're looking at it could actually be as wide or wider-

Table 2
Example of far analogy

1520	Tommy	like a garage door type of thing
1521	Todd	yeah push the button then it goes open

Table 3
Example of problem identification analogy

1204	Alan	in fact in some ways we should think about the fact it isn't even a pen
1205		because a pen you you'll always learn to write from left to right
1206		whether you're left handed or right handed
1207		so actually what you end up doing with left handed people is you smudge over over your work
1208		which is a problem
1209		but actually with this you're dragging it
1210		you're not pushing it are you
1211		most people will drag it

Table 4
Example of concept generation analogy

777	Alan	because the other thing that you use to make sure things are level that's come out in the sort of DIY world is these laser levellers and things like that
779		if you had like a little laser that made sure that it was level of some sort
780		erm you know the child can actually see a line
781		and that its at the right angle then

Table 5
Example of explanation analogy

213	Tommy	yeah this is a bit like photographic paper in a way
214		where you're erm developing what's on the paper
215		whereas here you're just enabling the bits you need to print
216		so here you're kind of getting in to normal text

Table 6
Example of function-finding analogy

1161	Tod	sort of like a like a could be like a finger puppe couldn't it
1162	Sandra	yeah cos wearing it like a finger puppet –
1163		the feel of it might be fun
1164	Tod	exactly so you can make you can make the footprints-

coded conceptually at the sentence/turn level, but tagged at the line level. Sentences/turns were coded as concept proposals any time a designer described a proposal for *how* to solve some design subproblem, where a design subproblem was defined as either (a) something the device (or a subsystem of it) has to do for the user (e.g., print, teach how to write, keep user's hands safe, make learning fun, make it harder to mess up, etc.) or (b) something the device or subsystem has to do to support or enable other functions (e.g., keep the print head level so that the print head mechanism can work). Defining concepts at the subproblem level provided external validity to the coding scheme, given the primary focus on concept generation, as concept generation in professional engineering practice routinely occurs following decomposition of an overall design problem into subproblems which are then addressed iteratively, sometimes in tandem (Ball, Evans, Dennis, & Ormerod, 1997; Ullman, 2002).

To avoid tagging of concept discussion lines as concept generation instances, only utterances that explicitly participated in a description of how a concept is meant to work were tagged as part of a concept; neither utterances evaluating concepts nor mere mentions of concepts (e.g., "that 'sheath idea' you mentioned earlier") were tagged as part of concepts unless they were embedded within a sentence or turn describing a concept. Through exhaustive triple coding, identification of concepts utterances was done at a high level of reliability; the intra-class correlation coefficient across the three coders was .88 (90% raw agreement).

To provide a further constraint on identification of concept utterances, coders also simultaneously proposed a segmentation for a coherent group of concept utterances into intact concepts and also proposed a pairing with one or more subproblems the concept was intended to address. Segmentation and pairing of concept utterances was then finalized by discussion during consensus meetings involving all three coders. In total, 217 unique concepts proposed for 42 subproblems were identified. Examples of subproblems included "keep the print head level," "specific application concept of product," "protect the print head," "power/energy saving," "user interface for controlling print options," "prevent overheating," "keep print head clean," "form of media," and "make device work for left-handed users."

Table 7 provides an example of a proposed design concept for the subproblem "keep the print head level." Due to the nature of the thermal printing technology, the thermal print head had to interface with the printing media within a strict range of angles in order for printing performance to be acceptable; however, the target market for the product concept, that is, young children between the ages of 5 and 7, was judged as particularly unlikely to hold pens and writing devices in stable ways. This subproblem was a major one discussed by the designers, and 35 distinct concepts were proposed for addressing it. The concept proposed in Table 7 was essentially a forcing function that would (via the

Table 7
Example of concept for "keep the print head level"

690	Alan	( ) can I just explore that last one in a little more detail
691		because when organisations- making sure they can only be correct in one way
692		so the design and shape of the thing so it can only be done in one way
693		and that's the correct way
694		because then there is less sort of learning to be done by the user

shape of the device) force a particular way of holding the device that would insure appropriate angles of contact.

2.2.4. Constructing conceptual search spaces

To characterize the designers' search patterns during concept generation, it was necessary to first characterize the search spaces. As functional distance of concepts within the search space was the focus, a functional similarity space for concepts within each subproblem space was constructed via pairwise comparison ratings of functional distance for each concept in each subproblem space. That is, within each subproblem space (e.g., "keep the print head level"), all concepts generated by the designers were rated for functional distance from all other concepts addressing the same subproblem. Two senior engineering undergraduate students (in mechanical and electrical engineering, respectively—both engineering subdisciplines highly relevant to the subproblems being solved by the designers) conducted the pairwise ratings of functional distance. These students were selected for their design experience and strong recommendations by engineering faculty with whom they had taken coursework.

Functional distance between pairs of concepts was rated on a scale ranging from 1 to 5. Distance coding was conceptualized as a degree of overlap rating, with the following anchor points: 1 = *very similar* (very substantial overlap, only trivial differences), 2 = *somewhat similar* (substantial overlap, but some non-trivial differences), 3 = *somewhat different* (some overlap, some differences), 4 = (little overlap, numerous differences), and 5 = *radically different* (very minimal/trivial overlap). Examples of 1 and 5 rated pairs are given in Table 8 (all concepts from the "keep the print head level" subproblem space).

The coding procedure was as follows. For each subproblem space, the two coders together first looked through the list of proposed concepts in the space and agreed upon an initial set of important points of contrast for comparing concepts. For example, for

Table 8
Example of concept pair ratings

Concept 1	Concept 2	Distance
No. 28: Laser mechanisms detect angle of contact and provide feedback to user	No. 29: Project multiple light points from device that converge when print head is at correct angle	1
No. 14: Device is toy with one or more wheels	No. 16: Put three ball bearings around print head to interface with media	1
No. 8: Use a different type of print head with more favorable angle tolerance	No. 86: Have a switch that controls print head action based on angular movement	5
No. 32: Add a dedicated feedback display that goes on user's wrist to give feedback on device angle	No. 84: Add disc around print head that restricts angle of contact with media	5

concepts proposed for the subproblem "keep the print head level," one point of contrast was "user versus device-centric approach" (e.g., user centric would be "give feedback to user and user adjusts accordingly," vs. "device has suspension system that adjusts for user action automatically"). Next, the coders independently generated functional distance ratings for all pairwise comparisons within the subproblem space, using the points of contrast as a guide for their judgments. The final step involved computations of inter-rater agreement and discussion of disagreements greater than 1-point difference; differences of 1 point were averaged to produce a final distance rating.

It should be noted that not all concepts entered into the analysis. Because the current analysis was focused on movement within a conceptual space, subproblems with less than three proposed concepts were excluded. The final set of concepts for analysis included 135 proposed concepts for nine major subproblems (see Table 9). Inter-rater reliability for this measure was excellent, with an intra-class correlation coefficient of .94 for ratings in the final set of concepts.

2.2.5. Constructing independent and dependent variables

2.2.5.1. Dependent variables: The primary dependent variable was distance from prior concepts. Two prior concept reference points were employed: (a) MIN FROM LAST 5—minimum distance from the prior five concepts and (b) JUST PRIOR—distance from the JUST PRIOR concept. The two reference points provide complementary views of the designers' patterns of conceptual search: MIN FROM LAST 5 provided a stricter measure of jumps through the conceptual space, as a given concept would have a high "distance from reference point" value, if it was substantially functionally different from all of the five concepts that immediately preceded it; JUST PRIOR provided a more circumscribed measure of jumps but one that might capture more localized movement in the conceptual space. For example, suppose the designers generated five concepts consecutively (C_1, C_2, C_3, C_4, and C_5). C_5 would receive a high "distance from reference point" value if it was substantially different from C_4, even if it was functionally similar to C_1, C_2, and C_3.

Although the ratings were technically obtained in an ordinal fashion, they are meant to approximate an interval scale, as is the case with the majority of Likert-type scales,

Table 9
Subproblems by number of concepts

Subproblem	No. Concepts
Keep the print head level	35
Specific application concept of product	35
Protect the print head	29
Acquiring print patterns	9
Powering the device	7
User interface for controlling print options	6
Varying print options available to user	6
Insure print head only fires when on media	5
Maintain appropriate surface area of contact between print head and media	3

which are frequently analyzed with ANOVAS, and results are most often very consistent with complementary analyses using non-parametric models. More important, we have direct evidence from our data that our distance measure behaves in a way that approximates an interval scale; the ratings for the three largest subproblem spaces (i.e., "keep the print head level," "specific application concept of product," and "protect the print head") closely approximate the triangle inequality (i.e., for any triangle, the sum of the lengths of any two sides must be greater than the length of the remaining side), an important property that must hold for distances in Euclidean space (which are interval scale; Beals, Krantz, & Tversky, 1968). Less than 1% of the triangles in the first two subproblem spaces, and less than 4% of triangles in the remaining subproblem space violate this inequality (most violations consist of the remaining side being within one point of the sum of the other two sides). For these reasons, we analyze our dependent variables as interval scales.

2.2.5.2. Independent variables: The primary independent variable was an ANALOGY BEFORE measure, which had two levels: (a) FAR ANALOGY, for concepts preceded by analogies that were both far *and* concept generating (function-finding analogies were included in this definition, as they served the purpose of generating new functional elements for a concept) and (b) *baseline,* for concepts not preceded by any far analogies (as defined in [a]). To thoroughly explore the space of possibilities for the effects of analogy, ANALOGY BEFORE was created at two different time windows: 10 and 5 lines prior to the concept onset. Number of lines rather than time per se was chosen as the segmentation unit of analysis because the focus was on information exchange and cognitive processes, which could happen at varying rates with respect to the passage of time per se. This range of time window sizes reflected our focus on relatively immediate effects of far analogies on concept generation.

The process of creating ANALOGY BEFORE for each of the time windows was identical and was as follows. For each concept, its initial onset in the transcript was identified. Next, the *n* lines prior to the onset were scanned to determine whether any of those lines contained at least part of an analogy/analogies, keeping separate track of *distance* and *purpose* of these analogy/analogies. With this information, concepts were classified into either the *baseline* or FAR ANALOGY groups; if a concept was preceded by an analogy that was not both far and concept generating, it was discarded. This allowed for a clean estimation of the effects of far concept-generating analogies on the conceptual search process. Hereafter, the term "far analogies" will be used as shorthand to refer to "far analogies used for concept generation." The number of concepts in each ANALOGY BEFORE level by reference point is shown in Table 10.

It should be noted that some concepts were preceded by multiple analogies. In these cases, the concept in question was classified based on the predominant distance and purpose of the analogies; more specifically, a concept was assigned to the FAR ANALOGY level if and only if the majority of the analogies (i.e., more than half) were far *and* either concept generating or function finding. In addition, given the naturalistic character of the data, the comparison to *baseline* is not to a standard "control" no input condition, but

Table 10
Number of concepts in each ANALOGY BEFORE condition at 10-line and 5-line windows at two different reference points for distance from prior concepts

	10-Line Window		5-Line Window	
Reference point	Baseline	FAR ANALOGY	Baseline	FAR ANALOGY
MIN FROM LAST 5	59	33	72	25
JUST PRIOR	81	38	95	30

Table 11
Mean (and standard error) functional distance for each ANALOGY BEFORE level at 10-line and 5-line windows, with two different reference points for distance from prior concepts

	10-line window		5-line window	
Reference point	Baseline	FAR ANALOGY	Baseline	FAR ANALOGY
MIN FROM LAST	2.1 (0.2)	2.0 (0.2)	2.1 (0.1)	1.9 (0.2)
JUST PRIOR	3.3 (0.2)	2.6 (0.2)	3.2 (0.1)	2.6 (0.2)

more precisely against functional distance of search when the designers were not using far analogies; other concept-generating strategies were more than likely being employed, such as reasoning from first principles and mutation of existing concepts (Gero & Maher, 1991; Ullman, 2002). That is, the study evaluates whether far analogies are particularly powerful, as the literature argues, rather than simply evaluating whether it has any effect at all.

2.3. Results

Four separate one-way ANOVAS were run for the two distance from reference point-dependent variables, two using MIN FROM LAST 5 as the dependent variable: (a) MIN FROM LAST 5 by ANALOGY BEFORE (10-line window) and (b) MIN FROM LAST 5 by ANALOGY BEFORE (5-line window); and two with JUST PRIOR as the dependent variable: (c) JUST PRIOR by ANALOGY BEFORE (10-line window) and (d) JUST PRIOR by ANALOGY BEFORE (5-line window). Distance from reference point means for each ANALOGY BEFORE level, for both 10-line and 5-line windows, are shown in Table 11.

2.3.1. MIN FROM LAST 5

2.3.1.1. 10-line window: There was no statistically significant main effect of ANALOGY BEFORE, $F(1, 95) = 0.36$, $p = .55$. Concepts were neither more nor less distant from their last five predecessors when preceded in the last 10 lines by far analogies versus baseline conditions, Cohen's $d = -0.06$ (95% confidence interval = −0.46 to 0.24).

2.3.1.2. 5-line window: There was no statistically significant main effect of ANALOGY BEFORE, $F(1, 90) = 0.08$, $p = .78$. Concepts were neither more nor less distant from their last five predecessors when preceded in the last five lines by far analogies versus baseline conditions, Cohen's $d = -0.14$ (95% confidence interval = −0.61 to 0.13).

2.3.2. *Just prior*

2.3.2.1. 10-line window: There was a statistically significant main effect of ANALOGY BEFORE measure, $F(1, 117) = 6.47$, $p = .01$, $\eta^2 = .05$. However, the nature of the effect was contrary to the initial hypothesis; concepts were less distant from their immediate predecessors when preceded in the last 10 lines by far analogies versus baseline conditions, Cohen's $d = -0.50$ (95% confidence interval = −0.90 to −0.21).

2.3.2.2. 5-line window: There was a statistically significant main effect of ANALOGY BEFORE, $F(1, 123) = 4.52$, $p = .04$, $\eta^2 = .04$. As with the 10-line window analysis, concepts were less functionally distant from their immediate predecessors when preceded in the last five lines by far analogies versus baseline conditions, Cohen's $d = -0.45$ (95% confidence interval = −0.90 to −0.18).

Fig. 1 illustrates the nature of the effect found in the ANOVAS with JUST PRIOR as the dependent variable. Each stacked bar presents percentage of concepts at each functional distance level in the two ANALOGY BEFORE levels (defined at the 5-line window). Attending first to the *baseline* bar, it is clear that jumps (distance from JUST PRIOR >3; the darker gray regions) are a common search step when designers were not using far analogies for concept generation, accounting for approximately half of all such concepts. Attending next to the FAR ANALOGY bar, the contrast with the baseline concepts in terms of relative distributions of search steps is clear; far analogies are followed by more hops (distance from JUST PRIOR ≤2; 50% of concepts) compared to baseline conditions (27% are hops). This pattern suggests that the biasing toward hops from immediate predecessors is not spurious (e.g., driven by a few outlier FAR ANALOGY-concept cases), but rather may be indicative of a general pattern of FAR ANALOGY's impact on creative concept generation, at least for these expert designers.

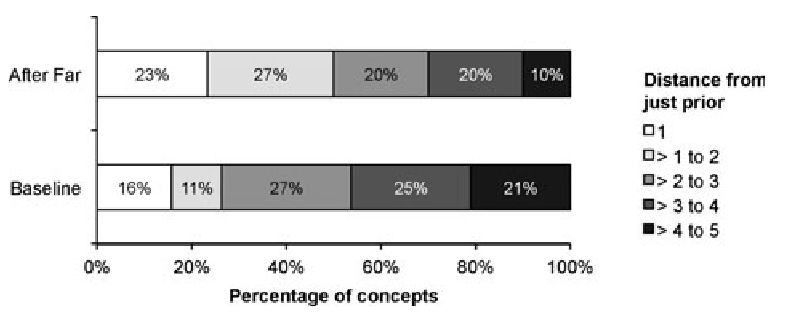

Fig. 1. Percentage of concepts at 5 distance from JUST PRIOR cutoff points, presented for baseline and FAR ANALOGY concepts, defined at the 5-line window.

2.4. Discussion

Overall, Study 1 found no support for the hypothesis that far analogies would lead to more jumps than hops, compared to baseline conditions; specifically, the analyses showed that the functional distance of proposed concepts from their immediate predecessors was *not* reliably greater when preceded by far analogies versus baseline. This result was robust across a range of time windows and measures. In fact, not only did functional distance from predecessors appear to be equivalent in the FAR ANALOGY versus *baseline* cases; when considering the distance of concepts from their immediate predecessors, FAR ANALOGY use was associated with conceptual moves that were *more* incremental than concept generation using other thought processes.

3. Study 2

Study 1's surprising counter-hypothesis findings raise questions surrounding the overall impact of far analogies on concept generation. The suppression effect on functional distance might be seen as evidence of fixation, in the sense of a decrease in the ability to generate concepts that are significantly different from ones already considered (Jansson & Smith, 1991; Smith, Ward, & Schumacher, 1993). This sort of fixation has been correlated with a decrease in the fluency of concept generation, another phenomenon that has been termed "fixation" due to the hypothesized importance of fluency for innovative outcomes (Guilford, 1950; Hennessey & Amabile, 2010; Runco, 2004; Shah, Vargas-Hernandez, & Smith, 2003; Terwiesch & Ulrich, 2009); for instance, increased fixation to example features during concept generation was associated with decreased levels of fluency (Chan et al., 2011), and the fluent generation of numerous concepts is empirically associated with the rate of generating novel, highly innovative concepts (Simonton, 1997). Thus, it is reasonable to ask whether the far analogies would also decrease fluency of concept generation.

Whether or not the far analogies decrease concept generation fluency has implications for the interpretation of Study 1's findings. If FAR ANALOGY use was associated with both suppressed functional distance of search *and* reduced concept generation fluency, it might be reasonable to suppose that the far analogies in this context were not productive (e.g., they were "fixating"). By contrast, if Study 2 did not yield evidence of suppression of concept generation fluency, Study 1's findings might be indicative not of the impact of unproductive far analogies but rather of a productive use of far analogies, focused on local idea exploration.

3.1. Overview

Given the *in vivo* and temporal nature of our data, we elected to examine the relationship between far analogies and concept generation fluency in terms of changes in the *probability* of generating concepts. Specifically, Study 2 examined whether FAR ANALOGY use was associated with a *lower* probability of concept generation relative to baseline levels. To address this question, a time-lagged logistic regression was employed; time lagged, because this analysis would estimate the change in concept generation probability at time t and $t + 1$ based on patterns of FAR ANALOGY use at time t, and logistic because the outcome variable was binary (i.e., did a designer generate a concept or not). This analysis assumed that (a) there was some baseline probability of a concept being generated in a given time slice and (b) a decrease in this probability as a function of the presence of a FAR ANALOGY in the current or previous time slice would suggest that the far analogies were reducing fluency of concept generation.

3.2. Methods

3.2.1. Creating blocks

The first step in the analysis was to segment the transcript into blocks for the time-lagged analysis. As similar trends were seen with block sizes of 10 and 5 lines in Study 1, and concepts were less rare than analogies, we selected a block size of five lines for this analysis to achieve a more favorable tradeoff between time window precision (estimating more immediate effects of FAR ANALOGY) and noise due to attrition (smaller time window leads to more attrition of measured phenomena).

Sets of five consecutive lines were chunked to create separate blocks. When a coherent cluster of analogy utterances occurred that contained at least one far concept-generating analogy (here, as with Study 1, this included both concept generation and function-finding analogies), it was marked as its own block, beginning from the start to the end of the analogy cluster. Subsequent sets of five consecutive lines continued to be clustered into separate blocks, until the next cluster of FAR ANALOGY utterances began (see Fig. 2 for a visual summary of the block creation strategy). Analogy onsets and offsets were used as boundary markers for blocks because the focus is on estimating the effects of analogy, which should be most directly shown when closely time locked to analogies. Because of

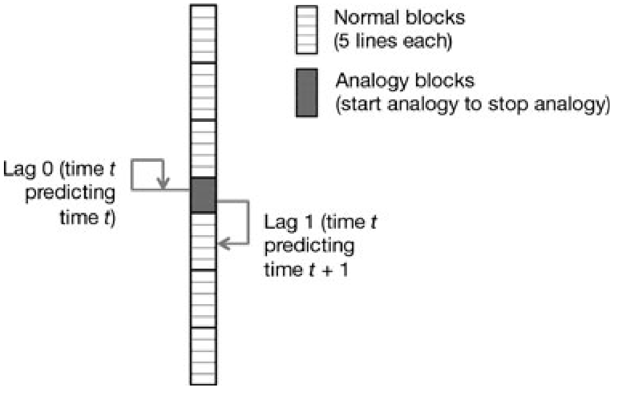

Fig. 2. Analogy-centered block creation strategy and time lags.

this analogy-centered block creation strategy, blocks immediately preceding analogy blocks were sometimes (fewer than 6% of blocks) less than five lines long.

This block creation strategy resulted in 97 analogy blocks and 843 non-analogy 5-line blocks. The reasons for the discrepancy between the number of analogy blocks and the number of unique analogies identified in the transcript (i.e., 147) are that (a) analogy clusters that did not contain far concept-generating analogies were treated as "normal" blocks, (b) analogies sometimes re-entered the conversation at later times, and (c) some analogy clusters were composed of more than one analogy (if they occurred in immediate succession). Analogy block lengths ranged from 1 line to 28 lines (M = 5.2, SD = 4.9), with most (88%) analogy blocks being 10 lines or less.

3.2.2. Independent and dependent variables

3.2.2.1. Independent variable: Similar to Study 1, the independent variable was FAR ANALOGY and had two levels: *yes*, if the block contained a FAR ANALOGY, and *no*, if it did not. Thus, as in Study 1, concept generation rates associated with far analogies were not compared with a traditional baseline, but rather with conditions in which other cognitive processes were being employed.

3.2.2.2. Dependent variable: The dependent variable, NEW CONCEPT, was a binary indicator for whether or not a NEW CONCEPT onset was present in the block (*yes* or *no*) regardless of functional distance to prior concepts; that is, a block was coded as "concept = yes" if and only if it contained an *onset* of a concept that was not mentioned in previous blocks. This ensured that the analysis would more cleanly reflect effects of far analogies on the *generation* (rather than elaboration) of concepts.

3.3. Results

Two separate time-lagged logistic regression models were estimated for lag 0 and lag 1 relationships between the FAR ANALOGY and NEW CONCEPT measures. The lag 0 model estimated the co-occurrence relationship between FAR ANALOGY at time t and NEW CONCEPT at time t; the lag 1 model estimated the relationship between FAR ANALOGY at time t and NEW CONCEPT at time $t + 1$ (i.e., in the next block; see Fig. 2 for a visual depiction of each time lag). Using only lags 0 and 1 focuses on immediate consequences that best fit the hypotheses under test and reduce the probability of finding spurious correlations from examining multiple lags.

The odds ratios for each lag are summarized in Table 12. The models did not show any decrease in concept generation as a function of FAR ANALOGY for either lag (nor did analyses using larger or smaller window sizes); on the contrary, FAR ANALOGY use was reliably associated with an *increase* in concept generation rate relative to baseline conditions, that is, when designers were engaging in processes other than using far analogies to generate concepts. For lag 0, the overall model was statistically significant, χ^2 (1, N = 938) = 8.02, p = .00, Negelkerke R^2 = .013, and the coefficient for the FAR ANALOGY predictor, β = .69, odds ratio = 1.99, indicated that FAR ANALOGY use was associated with

Table 12
Odds ratios by lag type for logistic regressions of NEW CONCEPT on FAR ANALOGY

	Odds ratio	95% CI	
		Lower Limit	Upper Limit
Lag 0	1.99**	1.26	3.15
Lag 1	1.88**	1.18	2.98

Note. **Denotes $p < .01$.

an approximately 100% increase in the odds of a concept being generated in the same block, relative to other processes the designer might otherwise be engaged in. This coefficient was statistically significant, Wald χ^2 (1) = 8.59, $p = .00$.

The lag 1 model estimates were very similar. The overall model was statistically significant, χ^2 (1, $N = 938$) = 6.63, $p = .01$, Negelkerke $R^2 = .011$, and the estimated coefficient for FAR ANALOGY, $\beta = .63$, odds ratio = 1.88, indicated that FAR ANALOGY use was associated with an approximately 88% increase in the odds of a concept being generated in the next block, relative to other processes the designer might otherwise be engaged in. This coefficient was statistically significant, Wald χ^2 (1) = 7.09, $p = .00$.

3.4. Discussion

Taken together, Study 2's results are not consistent with the hypothesis that the far analogies decreased fluency of concept generation. On the contrary, the positive odds ratios from the models indicated that the far analogies increased fluency of concept generation, even when compared to other concept-generating processes the designers might have been engaged in. These results suggest that Study 1's findings are not indicative of the impact of only unproductive far analogies, but they might be suggestive of far analogies spurring more functionally local conceptual search.

4. Study 3

In Study 3, we sought to provide additional tests of the potential relationship between FAR ANALOGY use and local conceptual search. One potential interpretation of the suppression of distance observed in Study 1, in tandem with the increased fluency found in Study 2, could be that the far analogies were being used to more deeply explore certain regions of the design space. Rietzschel and colleagues (Rietzschel, Nijstad, & Stroebe, 2007; Rietzschel, De Dreu, & Nijstad, 2009) have argued that novel concepts can often come from deep exploration within conceptual categories; because there are only a limited number of "conventional" concepts within categories, and initial forays into categories will tend to be superficial and be biased toward conventional ideas, extended exploration within categories can allow problem solvers to reach highly novel concepts within those categories. This conjecture is consistent with the findings of "extended effort" effects, where within an idea generation session, ideas generated later tend to be more novel than ideas generated earlier (Basadur & Thompson, 1986; Beaty & Silvia, 2012; Parnes, 1961). In the domain of design, Heylighen, Deisz, and Verstijnen (2007) showed that recombination and restructuring of elements within concepts for a design task (at the expense of lowering overall number of unique concepts) was correlated with more original concepts being produced.

It is possible that the designers were using the far analogies to generate variations on concepts that were different enough that they could continue to explore the design space more thoroughly. From an analogical retrieval perspective, too, one might expect to see such an effect of FAR ANALOGY use (i.e., generating more hops than jumps) in the context of relatively functionally coherent conceptual exploration; analogical comparison of two or more isomorphic or structurally very similar knowledge/solutions (as is the case with our data) can aid in the formation of an abstract schema through structural alignment (Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Loewenstein, Thompson, & Gentner, 1999), which can serve as a stronger base for retrieval of superficially dissimilar but structurally similar analogs from memory (Gentner, Loewenstein, Thompson, & Forbus, 2009; Kurtz & Loewenstein, 2007). However, this mechanism for increasing the probability of retrieving far analogies may also strongly favor retrieval of functionally very similar solutions, as structural similarity is the primary retrieval cue. Thus, observing a relatively coherent pattern of conceptual exploration JUST PRIOR to far analogies may help to explain why far analogies might be associated with incremental conceptual moves rather than jumps; far analogies may be more likely to be retrieved during an episode of exploration of variations on a common functional theme, and these far analogies are likely to be also functionally similar to the concepts being considered in that episode due to structural alignment.

4.1. Methods

To explore this potential explanation of the association between far analogies and reduced functional distance of search, we examined the concepts immediately preceding far analogy-to-concept pairs (i.e., far concept-generating or function-finding analogies), focusing on the distance of each concept from its immediate predecessor (i.e., its JUST PRIOR value, derived from Study 1). In building this sample of concepts, we screened out concepts that were not in the same subproblem space as the concept following the analogy, and concepts with predecessors in a different subproblem space. The final sample consisted of 57 concepts. The research question pursued was whether these concepts would, like the concepts preceded by far analogies, also be more likely to be functionally similar to their immediate predecessors, compared to baseline conditions. We used the 81 baseline concepts and 95 baseline concepts from Study 1 as the baseline benchmarks.

4.2. Results and discussion

Forty-one of the 57 concepts (72%) were themselves “hops” (i.e., distance of less than 3) from their immediate predecessors, although with far fewer “hops” among that set than in the set of concepts following far analogies (see Fig. 3 for a visual comparison of the distributions for functional distance from JUST PRIOR values for concepts in baseline conditions, preceded by far analogies, and preceding far analogies).

In statistical terms, the concepts that immediately preceded the FAR ANALOGY-concept pairs were less distant from their immediate predecessors compared to both 10-line window baseline concepts, Cohen's $d = -0.38$ (95% CI = -0.64 to -0.09), and 5-line window baseline concepts, Cohen's $d = -0.30$ (95% CI = -0.56 to -0.03). The former contrast was statistically significant at the conventional $\alpha = .05$ level using an independent samples t test, $t(136) = 2.17$, $p = .03$, while the latter contrast was marginally significant using the same α level with an independent samples t test, $t(150) = 1.78$, $p = .08$. These data suggest that the far analogies were often situated in a stream of relatively coherent conceptual exploration, where successive concepts (at least three in a row, two before the analogy, and one after) were variations of each other within a region of the design space.

5. General discussion

5.1. Summary and interpretation of findings

In summary, three studies were conducted to unpack in detail the effects of FAR ANALOGY use on conceptual search patterns in the naturalistic conversations of a real-world professional design team. Study 1 showed that the use of far concept-generating analogies was not associated with increased functional distance of proposed concepts from their predecessors. In fact, there was evidence that FAR ANALOGY use was temporally associated with *decreased* functional distance of search relative to immediate predecessors. Study 2 examined whether this effect was associated with an overall fixating effect, and showed that rather than decreasing the fluency of concept generation, far concept-generating analogies were associated with increased fluency, both during and after their use. This result

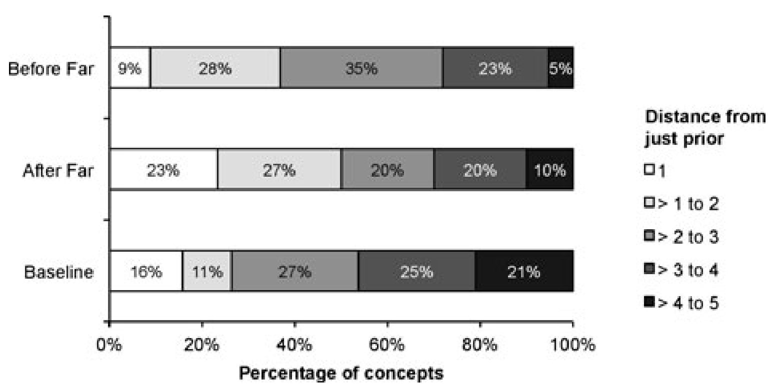

Fig. 3. Percentage of concepts at 5 distance from JUST PRIOR cutoff points, presented for baseline concepts, FAR ANALOGY concepts (defined at the 5-line window), and concepts immediately preceding far analogies.

helped to clarify the nature of FAR ANALOGY's impact on concept generation; rather than generally slowing down concept generation, the far analogies appeared to be used to keep the flow of concepts moving, with a special emphasis on generating functionally incremental steps in the conceptual space. Finally, in Study 3, we conducted an analysis of conceptual search patterns JUST PRIOR to FAR ANALOGY-concept pairs and found that concepts preceding FAR ANALOGY-concept pairs were also more likely to be functionally more similar to their predecessors (compared to baseline concepts). Altogether, the three studies suggest that, contrary to some previous accounts of creativity, far analogies may not lead to novel concepts via jumps in conceptual space; rather, far analogies may be embedded in and supportive of coherent streams of conceptual exploration, perhaps in support of a search for functionally novel concepts via deep search. To ground these quantitative observations and illustrate the effects found in the three studies, here we present two extracts from the transcripts that illustrate conceptual explorations involving far analogies.

In Table 13, the designers are searching for ways to protect the print head from being damaged by unexpected contact when the device is not printing, exploring a space of possible retractable covers for the print head. Two far analogies are employed to generate two distinct variations on this concept: Concept 61 involves a mechanism similar to a video tape flap with a rigid flap that opens to release the print head for use, while Concept 62 retains the core concept of a retractable cover, but using a slightly different mechanism, similar to a rolling garage door. Here, we see how both analogies were a source of concepts, and how the FAR ANALOGY to the garage door provided a way to further explore the space of retractable covers.

In Table 14, the designers are searching for concepts that address the subproblem of maintaining the optimal angle of contact between the print head and the media, given that the target users are young children who are unlikely to hold the printing device still to

Table 13
Example of progression in conceptual exploration involving FAR ANALOGY

Analogy: Video tape flap		
1516	Todd	I'm thinking of something a bit like erm the flap on a video tape
1517		*(pause)*
1518	Alan	uh-huh what the flap?
1519	Todd	yeah
Analogy: Garage door		
1520	Tommy	like a garage door type of thing
Concept 61		
1521	Todd	yeah push the button
1522		then it goes open
1523	Tommy	yeah
1524	Todd	but that's probably overly complicated
Concept 62		
1525	Rodney	garage door well it could be a roller
1526	Todd	a roller door

Table 14
Another example of progression in conceptual exploration involving FAR ANALOGY

Concept 24		
692	Alan	so the design and shape of the thing so it can only be done in one way
693		and that's the correct way
694		because then there is less sort of learning to be done by the user
Concept 25		
729	Alan	you could even have sort of some feedback
730		in terms of colour LEDs on the pen saying that he's done a good or she's done a good job or-laser levellers
Analogy: DIY		
777	Alan	because the other thing that you use to make sure things are level that's come out in the sort of DIY world is these laser levellers and things like that
Concept 28[a]		
779	Alan	if you had like a little laser that made sure that it was level of some sort
780		erm you know the child can actually see a line
781		and that its at the right angle then
782		because they can see that the line is right

Note. [a]Concept numbers not contiguous because concepts relating to other subproblems were discussed in between Concept 25 and the analogy.

achieve that angle of contact without some help. The designer proposes Concept 24, which involves designing the shape of the device such that it forces the user to hold it in the "correct" way (i.e., in a way that preserves the optimum angle of contact between the print head and the media). Concept 25 takes the idea of feedback in a slightly different direction and proposes giving visual (as opposed to simply tactile) feedback to the user to guide interactions. This delineates a region of the design space with the general approach of providing perceptible feedback and sets the stage for a FAR ANALOGY to laser levelers, used among DIY enthusiasts to make sure they are following appropriate angles for various construction and re-pair tasks (e.g., laying tiles, constructing shelves, etc.). Again, as in Table 13, this analogy was a direct source of Concept 28, which spurred further exploration within that region of the design space by changing the way the feedback would be provided to the user, while retaining key functional features from Concept 25.

Together, these extracts illustrate how far analogies were a significant source of concepts that tended to be embedded in and supportive of continued explorations in particular functional regions of the design space, rather than large functional jumps.

5.2. *Caveats*

Some caveats should be mentioned before discussing the broader implications of this work. First, the present empirical approach involved a tradeoff between external and internal validity. While the naturalistic character of the data and the fact that the designers are real-world professionals lend external validity to the findings, it should be noted that the findings are correlational in nature, and tight experimental control of potential confounding variables was not possible. Nevertheless, our data have several mitigating factors that lend strength to the internal validity of the findings, namely the high reliability of the measures (e.g., $k > .8$, ICC $> .9$ for analogy distance and concept functional distance, respectively), the descriptive analysis of the frequency distributions of functional distance of concepts addressed by far analogies versus not, the analysis of the analogy-concept extracts from the transcript in the discussion, and the examination of *temporal* order.

A related caveat has to do with the tradeoff between depth and breadth; the data collection, coding, and analytic methods employed in the present work, while affording highly detailed looks at the temporal interplay between analogy use and concept generation, are highly resource intensive, making comparisons across multiple expert datasets difficult. From one perspective, the sample size of the three studies was essentially $N = 1$, given that only one team was studied. Nevertheless, the high external validity of the data does provide some initial confidence that the observed interplay between far analogies and conceptual search patterns is as likely to generalize to other real-world contexts as studies conducted in the laboratory. Further, the team worked on many different subproblems, and thus the observed pattern is unlikely to be driven by characteristics of a single problem; many laboratory studies employ many participants, but all participants often solve a particular problem.

The restriction to one team also precludes our ability to relate the observed patterns directly to final creative outcomes; thus, only descriptive (*not* prescriptive) inferences are supported by our data. Our data are silent on whether the patterns of relationships between far analogies and conceptual search patterns are low-performance or high-performance creative concept generation strategies. Yet the designers were experienced, professional designers at a firm known for innovation, suggesting that the patterns observed may represent an expert concept generation strategy involving far analogies.

Finally, some might be concerned about our reliance on verbal reports as data. Such methods can suffer from loss of signal; however, the critical question is whether our loss of signal is *systematic* in a way that undermines our analyses and inferences. For instance, our choice to measure verbally expressed analogy precludes measurement of "implicit" analogy (i.e., mappings that occur below/without conscious awareness); however, given that most theories of analogy assume the central mapping process occurs in working memory, we do not believe that these implicit mappings are actually analogy at work (see, e.g., Schunn & Dunbar, 1996). With respect to explicit but not verbalized analogies, we believe the interactive nature of the design meetings helps to mitigate concerns about missing such analogies. Transcripts of collaborative discussions can be thought of as approximating the level of explication of thought in individual verbal protocols, as the collaborators have an incentive to provide common ground for collaborative problem solving, particularly given the multidisciplinary context. There could also be a systematic loss of signal biasing against truly far analogies due to social inhibition; however, we believe this is not present in our data, as the designers were given standard brainstorming instructions to encourage wild ideas, had been working together for many years, and many outlandish things were said in the meeting (e.g., evil emperor from Star Wars with lightning bolts shooting out of nose, joking that they should teach left-handed children to be right handed by hitting them with a cane). For these reasons, we accept that there is some loss of signal in our analogy measure, but we do not believe that there is significant or systematic loss of signal that precludes our ability to draw useful inferences from the data.

5.3. Future directions

We now note some key future directions of this work. First, there is the issue of generalizability across design situations. It is possible that FAR ANALOGY-generated jumps may only occur in certain design situations. Perkins (1994, 1997) has described a potential "isolation problem" in creative problem spaces, where innovative concepts are bounded in the space by wildernesses of no promise. In these situations, incremental search may lead to an impasse, as there is no incremental path into the location of the innovative concept that avoids going through highly unpromising options. It may be that large jumps into these isolated regions of promise might be facilitated by highly functionally distant analogies, perhaps sparked by external stimulations. This notion is consistent with the literature on incubation and "prepared mind" effects, where creative problem solvers overcome impasses in their problem solving by unexpectedly encountering potentially relevant ideas in their environment after having set their problem aside (Christensen & Schunn, 2005; Seifert et al., 1995; Tseng et al., 2008). These ideas suggest that impasses may be a prerequisite for observing jumps supported by analogy.

The data we have do not allow us to speak directly to this issue, as we did not measure the occurrence of impasses; we did have an indirect measure of impasses (i.e., expressed uncertainty in their speech; for more information on the measure, see Ball & Christensen, 2009) but found no measurable difference in uncertainty levels between problems addressed by far analogies versus not. It is possible that the lack of increased uncertainty for problems addressed by analogy indicates a lack of impasses and therefore reduced likelihood of or need for large jumps; however, jumps did occur for problems not addressed by analogy, which had comparable levels of uncertainty (or lack thereof). Thus, our data are inconclusive regarding any potential variations in the relationship between far analogies and conceptual jumps as a function of impasses. Follow-up work may explore this issue further by creating impasse-likely and impasse-unlikely design situations and comparing the impact of far analogies on conceptual search patterns across those settings.

Further, there is the issue of self-generated nature of the far analogies in this data; that is, with just a few exceptions, most of the analogies were retrieved from the designers' memories. The few analogies that might have been retrieved from external sources were those generated prior to the first meeting; a meeting brief was sent around to the team prior to the first meeting, advising the designers of the major issues to be discussed in the two meetings (e.g., the angle problem, protecting the print head), and instructing the team members to bring to the meeting products or designs that have to glide smoothly over contours, to help kick-start concept generation for the angle problem. The primarily self-generated character of the analogies stands in contrast to the externally given analogies in many of the prior studies of analogy in design. In light of this, one possible explanation of the local/incremental character of conceptual search supported by FAR ANALOGY might be that many of the far analogies were insufficiently “far” from previously considered concepts.

That is, notwithstanding the documented capacity of people to retrieve far analogies from long-term memory in naturalistic settings (Blanchette & Dunbar, 2000; Dunbar, 2001), it is possible that, given the computational constraints of analogy (e.g., preferring systematic matches, one-to-one mappings; Gentner, 1983), the strong influence of surface similarity on retrieval (Forbus, Gentner, & Law, 1994; Gentner & Landers, 1985; Gentner, Rattermann, & Forbus, 1993; Keane, 1987; Rattermann & Gentner, 1987; Reeves & Weisberg, 1994; Ross, 1987), and the associative character of memory (Collins & Loftus, 1975; Raaijmakers & Shiffrin, 1981), designers might not be able to retrieve from memory other concepts that solve similar subproblems in very different ways, especially if these concepts are embedded within designs or products with very different overall functionality. Further, as noted earlier, most of the far concept-generating analogies appeared to have been retrieved within an episode of relatively coherent functional exploration, providing further constraints on the range of functional distance the designers could explore using analogical retrieval.

Different effects of analogy on conceptual search patterns might be observed with externally provided analogous sources that are highly distant functionally. Perhaps very far (even bordering on “random”) analogical stimuli from external sources are needed to truly support large conceptual jumps into novel search space territory. It is worth mentioning, however, that the current empirical support for the benefits of “random” analogies is mixed at best (for a recent review, see Christensen & Schunn, 2009a). There are also potentially important interactions between problem space structure and analogical source. It may be that far analogies retrieved from memory generally support increased fluency of search but enable jumps out of local maxima only in impasse situations, and perhaps only if they are “far enough” (e.g., from “random” external sources). Future work should explore these novel hypotheses.

5.4. Broader implications

We conclude by noting some broader implications of the work for understanding innovation in general from a cognitive standpoint. One potential insight might be an elevation of the importance of incremental/iterative development of concepts as a pathway to novel concepts. Insofar as far analogies in the concept generation process are associated with more innovative outcomes, we might infer from the present data that incremental accumulation of many small insights is at least as likely to lead to innovative outcomes as direct generation of very novel concepts. The history of innovation contains accounts of such “incremental” accumulations that culminated in innovative breakthroughs; one striking example is the invention of the steam engine by James Watt, which was powered in large part by a crucial addition of a steam condenser (for increased efficiency of the heating/cooling mechanism of the metal cylinder in the steam engine) to Newcomen’s “atmospheric engine”—this relatively small addition proved to be such a difference maker that

James Watt is often credited for the invention of the steam engine. Detailed *in vivo* and cognitive-historical accounts of innovation have also highlighted this very incremental pathway to highly innovative outcomes (Carlson & Gorman, 1990; Gorman, 1997; Weisberg, 2009).

Another potential implication might be a rethinking of the impact of analogical distance. If the cognitive mechanisms by which far analogies inspire innovation are shown to be very similar (or identical to) the inspirational mechanisms of near analogies (e.g., increased fluency), this might provide some motivation to question the fundamental distinction between far and near analogies in terms of their potential for supporting innovation. It may be that it is not analogical distance from one's problem per se that matters, as Perkins (1983) and Weisberg (2009) argue, but other considerations, such as the similarity of the analogical source to one's currently considered concepts, or the relationship of the analogy to other considered analogies (e.g., conceptual diversity of sources considered; Mumford, Baughman, & Sager, 2003; Taylor & Greve, 2006). Certainly, much more theoretical and empirical work is needed to evaluate whether this theoretical questioning is warranted.

Overall, the present work highlights the important and complementary role of detailed *in vivo* studies of cognition for a complete cognitive science of innovation; just as protocol analyses of online problem solving yielded invaluable insights that constrained theories of problem solving and aided in suggesting hypotheses for and guiding interpretations of experimental studies, so *in vivo* studies of the innovation process can continue to complement experimental data from input–output studies and inform more complete theories of the cognitive processes that lead to innovation.

Acknowledgments

This research was supported in part by National Science Foundation Grant #CMMI-0855293. We are grateful to Bo Christensen and Linden Ball for sharing this dataset with us and providing feedback on the conceptualization of the study; Nick Rassler, Sophia Bender, Adam Sparacino, Matt Oriolo, Oreste Scioscia, and Stephen Denninger for help with data processing and coding; and to Timothy Nokes-Malach, Kenneth Kotovsky, and Jonathan Cagan for comments and discussion on earlier drafts of this article.

References

Ball, L. J., & Christensen, B. T. (2009). Analogical reasoning and mental simulation in design: Two strategies linked to uncertainty resolution. *Design Studies*, *30*(2), 169–186.

Ball, L. J., Evans, J. S., Dennis, I., & Ormerod, T. C. (1997). Problem-solving strategies and expertise in engineering design. *Thinking and Reasoning*, *3*(4), 247–270.

Basadur, M., & Thompson, R. (1986). Usefulness of the ideation principle of extended effort in real world professional and managerial creative problem solving. *The Journal of Creative Behavior*, *20*(1), 23–34.

Beals, R., Krantz, D. H., & Tversky, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, *75*(2), 127–142.

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts*, *6* (4), 309–319.

Blanchette, I., & Dunbar, K. N. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition*, *28*(1), 108–124.

Blanchette, I., & Dunbar, K. N. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition*, *29*(5), 730–735.

Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. New York: Routledge.

Burt, R. S. (2004). Structural holes and good ideas. *The American Journal of Sociology*, *110*(2), 349–399.

Carlson, W. B., & Gorman, M. E. (1990). Understanding invention as a cognitive process: The case of Thomas Edison and early motion pictures, 1888-91. *Social Studies of Science*, *20*(3), 387–430.

Chakrabarti, A. (2006). *Defining and supporting design creativity*. In Design 2006: The 9th International Design Conference. Dubrovnik, Croatia: The Design Society.

Chan, J., Fu, K., Schunn, C. D., Cagan, J., Wood, K. L., & Kotovsky, K. (2011). On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of Mechanical Design*, *133*, 081004.

Chiu, I., & Shu, H. (2012). Investigating effects of oppositely related semantic stimuli on design concept creativity. *Journal of Engineering Design*, *23*(4), 271–296. doi:10.1080/09544828.2011.603298.

Christensen, B. T., & Schunn, C. D. (2005). Spontaneous access and analogical incubation effects. *Creativity Research Journal*, *17*(2), 207–220.

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

Christensen, B. T., & Schunn, C. D. (2009a). "Putting blinkers on a blind man": Providing cognitive support for creative processes with environmental cues. In A. B. Markman & K. L. Wood (Eds.), *Tools for innovation* (pp. 48–74). New York: Oxford University Press.

Christensen, B. T., & Schunn, C. D. (2009b). The role and impact of mental simulation in design. *Applied Cognitive Psychology*, *23*(3), 327–344.

Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, *12*(4), 563–586.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Dahl, D. W., & Moreau, P. (2002). The Influence and value of analogical thinking during new product ideation. *Journal of Marketing Research*, *39*(1), 47–60.

Dewar, R. D., & Dutton, J. E. (1986). The adoption of radical and incremental innovations: An empirical analysis. *Management Science*, *32*(11), 1422–1433.

Dunbar, K. N. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 461–493). Washington, DC: American Psychological Association.

Dunbar, K. N. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. In D. Gentner, K. J. Holyoak, & B. K. Kokinov (Eds.), *The analogical mind* (pp. 313–334). Cambridge, MA: MIT Press.

Dym, C. L. (1994). *Engineering design: A synthesis of views*. New York: Cambridge University Press.

Forbus, K. D., Gentner, D., & Law, K. (1994). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–205.

French, M. (1988). *Invention and evolution: Design in nature and engineering*. Cambridge, UK: Cambridge University Press.

French, R. M. (2002). The computational modeling of analogy-making. *Trends in Cognitive Sciences*, *6*(5), 200–205.

Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.

Gentner, D., Brem, S., Ferguson, R. W., Wolff, P., Markman, A. B., & Forbus, K. D. (1997). Analogy and creativity in the works of Johannes Kepler. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 403–459). Washington, DC: American Psychological Association.

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *WIREs Cognitive Science*, *2*, 266–276.

Gentner, D., & Landers, R. (1985). *Analogical reminding: A good match is hard to find*. In Proceedings of the 1985 International Conference on Cybernetics and Society. Tucson, AZ: IEEE.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–405.

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science*, *33*(8), 1343–1382.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, *25*(4), 524–575.

Gero, J. S. (2000). Computational models of innovative and creative design processes. *Technological Forecasting and Social Change*, *64*, 183–196.

Gero, J. S., & Kazakov, V. (1998). Using analogy to extend the behaviour state space in design. In J. S. Gero, & M. L. Maher (Eds.), *Computational models of creative design IV* (pp. 113–143). Sydney, Australia: Key Centre of Design Computing and Cognition, University of Sydney.

Gero, J. S., & Maher, M. L. (1991). *Mutation and analogy to support creativity in computer-aided design*. In proceedings of CAAD futures '91. Zurich, Switzerland: Elsevier.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15* (1), 1–38.

Goel, V. (1995). *Sketches of thought*. Cambridge, MA: MIT Press.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science*, *16*, 395–429.

Goldschmidt, G. (2001). Visual analogy: A strategy for design reasoning and learning. In C. M. Eastman, W. M. McCracken, & W. C. Newstetter (Eds.), *Design knowing and learning: Cognition in design education* (pp. 199–220). Amsterdam: Elsevier.

Gonçalves, M., Cardoso, C., & Badke-Schaub, P. (2013). Inspiration peak: Exploring the semantic distance between design problem and textual inspirational stimuli. *International Journal of Design Creativity and Innovation*, *1*(4), 214–232.

Gordon, W. J. J. (1961). *Synectics: The development of creative capacity*. New York: Harper and Brothers.

Gorman, M. E. (1997). Mind in the world: Cognition and practice in the invention of the telephone. *Social Studies of Science*, *27*(4), 583–624.

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454.

Hacco, E., & Shu, L. H. (2002). *Biomimetic concept generation applied to design for remanufacture*. 2002 ASME Design Engineering Technology Conference and Company and Information in Engineering Conference. Montreal, Quebec, Canada.

Hargadon, A. B. (2002). Brokering knowledge: Linking learning and innovation. *Research in Organizational Behavior*, *24*, 41–85.

Hender, J. M., Dean, D. L., Rodgers, T. L., & Jay, F. F. (2002). An examination of the impact of stimuli type and GSS structure on creativity: Brainstorming versus non-brainstorming techniques in a GSS environment. *Journal of Management Information Systems*, *18*(4), 59–85.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, *61*, 569–598.

Hey, J., Linsey, J., Agogino, A. M., & Wood, K. L. (2008). Analogies and metaphors in creative design. *International Journal of Engineering Education*, *24*(2), 12.

Heylighen, A., Deisz, P., & Verstijnen, I. M. (2007). Less is more original? *Design Studies*, *28*(5), 499–512.

Hofstadter, D. R. (2001). Epilogue: Analogy as the core of cognition. In D. Gentner, K. J. Holyoak, & B. K. Kokinov (Eds.), *The analogical mind* (pp. 499–538). Cambridge, MA: MIT Press.

Holyoak, K. J., & Thagard, P. (1996). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Huh, Y., & Kim, M. S. (2012). Study on creativity of game graphics. *Embedded and Multimedia Computing Technology and Service*, *181*, 339–346.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427–466.

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies*, *12*(1), 3–11.

Jones, B. F. (2009). The burden of knowledge and the "death of the Renaissance man": Is innovation getting harder? *Review of Economic Studies*, *76*(1), 283–317.

Keane, M. (1987). On retrieving analogues when solving problems. *The Quarterly Journal of Experimental Psychology*, *39*(1), 29–41.

Koestler, A. (1964). *The act of creation*. Oxford, England: Macmillan.

Kornish, L. J., & Ulrich, K. T. (2014). The importance of the raw idea in innovation: Testing the sow's ear hypothesis. *Journal of Marketing Research*, *51*(1), 14–26.

Kurtz, K. J., & Loewenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition*, *35*(2), 334–341.

Linsey, J., Murphy, J., Laux, J., Markman, A. B., & Wood, K. L. (2009). Supporting innovation by promoting analogical reasoning. In A. B. Markman, & K. L. Wood (Eds.), *Tools for innovation* (pp. 85–103). New York: Oxford University Press.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotation. *Psychonomic Bulletin & Review*, *6*(4), 586–597.

Malaga, R. A. (2000). The effect of stimulus modes and associative distance in individual creativity support systems. *Decision Support Systems*, *29*(2), 125–141.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220–232.

Mumford, M. D., Baughman, W. A., & Sager, C. E. (2003). Picking the right material: Cognitive processing skills and their role in creative thought. In M. A. Runco (Ed.), *Critical creative processes: Perspectives on creativity research* (pp. 19–68). Creskill, NJ: Hampton Press.

Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science* (Vol. 15, pp. 3–45). Minneapolis, MN: University of Minnesota Press.

Oppenheimer, R. (1956). Analogy in science. *American Psychologist*, *11*(3), 127–135.

Paletz, S. B. F., & Schunn, C. D. (2010). A social-cognitive framework of multidisciplinary team innovation. *Topics in Cognitive Science*, *2*(1), 73–95.

Parnes, S. J. (1961). Effects of extended effort in creative problem solving. *Journal of Educational Psychology*, *52*(3), 117–122.

Perkins, D. N. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.

Perkins, D. N. (1983). Novel remote analogies seldom contribute to discovery. *The Journal of Creative Behavior*, *17*(4), 223–239.

Perkins, D. N. (1994). Creativity: Beyond the Darwinian paradigm. In M. A. Boden (Ed.), *Dimensions of creativity* (pp. 119–142). Cambridge, MA: MIT Press.

Perkins, D. N. (1997). Creativity's camel: The role of analogy in invention. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 523–538). Washington, DC: American Psychological Association.

Poze, T. (1983). Analogical connections: The essence of creativity. *The Journal of Creative Behavior*, *17*(4), 240–258.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134.

Rattermann, M. J., & Gentner, D. (1987). *Analogy and similarity: Determinants of accessibility and inferential soundness*. In Proceedings of the Ninth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum.

Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, *115*(3), 381–400.

Rietzschel, E. F., De Dreu, C. K. W., & Nijstad, B. A. (2009). What are we talking about, when we talk about creativity? Group creativity as a multifaceted, multistage phenomenon. In E. A. Mannix, M. A. Neale, & J. A. Goncalo (Eds.), *Creativity in groups* (Vol. 12, pp. 1–27). Bingley, UK: Emerald.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology*, *43*(6), 933–946.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 629–639.

Ruef, M. (2002). Strong ties, weak ties and islands: Structural and cultural predictors of organizational innovation. *Industrial and Corporate Change*, *11*(3), 427–449.

Runco, M. A. (2004). Creativity. *Annual Review of Psychology*, *55*, 657–687.

Sawyer, K. (2007). *Group genius: The creative power of collaboration*. New York: Basic Books.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*, 182–203.

Schunn, C. D., & Dunbar, K. N. (1996). Priming, analogy, and awareness in complex reasoning. *Memory and Cognition*, *24*(3), 271–284.

Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1995). Demystification of cognitive insight- Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (Vol. 124, pp. 65–124). Cambridge, MA: MIT Press.

Shah, J. J., Vargas-Hernandez, N., & Smith, S. M. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, *24*(2), 111–134.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press.

Simonton, D. K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, *104*(1), 66–89.

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*(6), 837–845.

Sood, A., & Tellis, G. J. (2005). Technological evolution and radical innovation. *Journal of Marketing*, *69*, 152–168.

Taylor, A., & Greve, H. R. (2006). Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Academy of Management Journal*, *49*(4), 723–740.

Terwiesch, C., & Ulrich, K. T. (2009). *Innovation tournaments: Creating and selecting exceptional opportunities*. Boston, MA: Harvard Business Press.

Tortoriello, M., & Krackhardt, D. (2010). Activating cross-boundary knowledge: The role of Simmelian ties in the generation of innovations. *Academy of Management Journal*, *53*(1), 167–181.

Tseng, I., Moss, J., Cagan, J., & Kotovsky, K. (2008). The role of timing and analogical similarity in the stimulation of idea generation in design. *Design Studies*, *29*(3), 203–221.

Ullman, D. (2002). *The mechanical design process* (3rd ed.). New York: McGraw-Hill Professional.

Vargas-Hernandez, N., Shah, J. J., & Smith, S. M. (2010). Understanding design ideation mechanisms through multilevel aligned empirical studies. *Design Studies*, *31*(4), 382–410.

Walters, H. (2013). Since the TED Talk: Michael Pritchard on how he's helping end water poverty in Malaysia [Blog post]. Available at: http://blog.ted.com/2013/07/10/since-the-ted-talk-michael-pritchard-on-how-hes-helping-end-water-poverty-in-malaysia/. Accessed July 12, 2013.

Ward, T. B. (1998). Analogical distance and purpose in creative thought: Mental leaps versus mental hops. In K. J. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences* (pp. 221–230). Sofia, Bulgaria: New Bulgarian University.

Weisberg, R. W. (2009). On "out-of-the-box" thinking in creativity. In A. Markman & K. Wood (Eds.), *Tools for innovation* (pp. 23–47). New York: Oxford University Press.

Wilson, J. O., Rosen, D., Nelson, B. A., & Yen, J. (2010). The effects of biological examples in idea generation. *Design Studies*, *31*(2), 169–186.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Joel Chan
University of Pittsburgh,
LRDC Room 823, 3939 O'Hara Street,
Pittsburgh, PA 15260
e-mail: joc59@pitt.edu

Katherine Fu
Carnegie Mellon University,
5000 Forbes Avenue,
Pittsburgh, PA 15213
e-mail: kfu@andrew.cmu.edu

Christian Schunn
University of Pittsburgh,
LRDC Room 821, 3939 O'Hara Street,
Pittsburgh, PA 15260
e-mail: schunn@pitt.edu

Jonathan Cagan
Carnegie Mellon University,
Scaife Hall 419, 5000 Forbes Avenue,
Pittsburgh, PA 15213
e-mail: jcag@andrew.cmu.edu

Kristin Wood
University of Texas-Austin,
1 University Station, ETC 4.146B,
M/C C2200,
Austin, TX 78712-1063
e-mail: ood@mail.utexas.edu

Kenneth Kotovsky
Carnegie Mellon University,
Baker Hall 342F,
Pittsburgh, PA 15213
e-mail: kotovsky@andrew.cmu.edu

On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples

Drawing inspiration from examples by analogy can be a powerful tool for innovative design during conceptual ideation but also carries the risk of negative design outcomes (e.g., design fixation), depending on key properties of examples. Understanding these properties is critical for effectively harnessing the power of analogy. The current research explores how variations in analogical distance, commonness, and representation modality influence the effects of examples on conceptual ideation. Senior-level engineering students generated solution concepts for an engineering design problem with or without provided examples drawn from the U.S. Patent database. Examples were crossed by analogical distance (near-field vs. far-field), commonness (more vs. less-common), and modality (picture vs. text). A control group that received no examples was included for comparison. Effects were examined on a mixture of ideation process and product variables. Our results show positive effects of far-field and less-common examples on novelty and variability in quality of solution concepts. These effects are not modulated by modality. However, detailed analyses of process variables suggest divergent inspiration pathways for far-field vs. less-common examples. Additionally, the combination of far-field, less-common examples resulted in more novel concepts than in the control group. These findings suggest guidelines for the effective design and implementation of design-by-analogy methods, particularly a focus on far-field, less-common examples during the ideation process. [DOI: 10.1115/1.4004396]

Keywords: *design cognition, design methods, conceptual design, innovation, analogy*

1 Introduction

Innovation, defined as the capacity to generate ideas or products that are both novel and useful, is a critical component of successful design in today's economy [1,2]. A number of investigators have argued that innovation can be best managed in the "fuzzy front end" of the design process [3,4], notably in the ideation phase, where concepts are created either intuitively or through systematic processes. While many approaches exist to create ideas and concepts as part of ideation, the search for and use of analogies have been shown to be quite powerful [5–8]. Analogy is a mapping of knowledge from one domain to another enabled by a supporting system of relations or representations between situations [9]. This process of comparison between situations fosters new inferences and promotes construing problems in new insightful ways. This process likewise is dependent on how the problem is represented, encouraging multiple representations to more fully enable analogical reasoning [10,11]. As an illustrative example, the design concept for the bipolar plate of a fuel cell could be usefully informed by analogy to a plant leaf due to its similarity in functionality. The most significant functions affecting the current generation capability of a bipolar plate are "distribute fluid," "guide fluid," and "disperse fluid." The plant leaf possesses a similar function chain, where the veins and lamina perform the functions. As a result of this analogy, the bipolar plate flow field can be designed to mimic the structure of a leaf [10,11].

Design-by-analogy is clearly a powerful tool in the conceptual design process, and a number of methods have been developed to harness its power, such as Synectics [12]—group design through analogy types; French's work on inspiration from nature [13]; Biomimetic concept generation [14]—a systematic tool to index biological phenomena that links to textbook information; and analogous design using the Function and Flow Basis [15,16]—analogous and nonobvious product exploration using the functional and flow basis. However, fundamental questions surround the proper use of design-by-analogy methods. Most critical, and the problems that are the focus in our work, are what should one analogize over, and what reasoning modalities and associated representations make innovative design-by-analogy more likely?

While these questions have remained largely unanswered in specific knowledge domains such as engineering design, there is related research literature in the domain of psychological studies of creativity, reasoning, and problem solving. In what follows, we review the relevant literature that motivate our present hypotheses, describe the methods and findings of our cognitive study, and then discuss the insights and implications of our work.

2 Background

2.1 Analogical Distance of Example Designs. One key variable of interest with respect to the question of what one should analogize over is analogical distance. This variable can be

Contributed by the Design Education Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received December 20, 2010; final manuscript received June 7, 2011; published online August 1, 2011. Assoc. Editor: Janis Terpenny.

conceptualized as ranging over a continuum from far-field (from a different problem domain) to near-field (from the same or very similar problem domain), where analogies closer to the far-field end point share little or no surface features with the target domain, while analogies closer to the near-field end point share a significant number of surface features. The potential for creative insights seems clearest when the two domains being compared are very different on the surface [17]. Classic accounts of creative discoveries and inventions often highlight the potential of far-field analogies for creative insights, including George Mestral's invention of Velcro via analogy to burdock root seeds, and Niels Bohr's discovery of the structure of atoms via analogy to the solar system. Empirical work has also supported a link between far-field analogies and innovative outcomes. For instance, it has been shown that the number of far-field analogies used by designers during ideation is positively related to the originality of proposed solutions, as rated by a sample of potential customers [18]. Further, exposure to surface dissimilar design examples increases idea novelty relative to using no examples, and exposure to surface similar examples decreases the variety of ideas generated relative to surface dissimilar examples [19].

On the other hand, far-field analogies can be difficult to retrieve from memory [20] or notice as relevant to one's target problem [5]. In addition, some investigators have disputed the privileged role of far-field analogies in prominent inventions and discoveries [21,22]. As such, it is an open question whether far-field analogies are always beneficial to the design process. One way to tease apart possible ways in which far-field and near-field analogies might help or hinder designers is to use multiple measures of ideation processes, including novelty and variety of ideas, as well as average quality and variance in idea quality. An initial testable hypothesis is that providing far-field examples would allow one to generate more novel ideas relative to near-field or no examples.

2.2 Commonness of Example Designs. Another potential variable of interest is the commonness of example designs (i.e., how common the designs are found in designers' worlds). The commonness of the example design in its respective design space increases the probability that a designer would have had prior exposure and/or experience with the design. Psychologically, the commonness of an example design is related to the degree to which it activates relevant prior knowledge of a designer. This knowledge can come from exposure to instances (since designed objects exist in the world), or from deliberately structured experiences, such as in engineering coursework or in the course of professional design [23]. The psychological literature on creativity and problem solving suggests that prior experience with an artifact might influence one's ability to flexibly re-represent and use it and combine it with other concepts in a novel fashion. Take for instance, Duncker's [24] classic candle problem, where the task is to fix a lighted candle on a wall in such a way that the candle wax will not drip onto a table below, and the given materials are a candle, a book of matches, and a box of thumb-tacks. A correct solution involves emptying the box of tacks and using it as a platform for the candle; however, this solution eludes most solvers because it requires recognizing an unconventional use of the box as a platform. In fact, when the box is presented to solvers empty, with the tacks beside it, solvers are much more likely to find the unconventional solution [25]. Similarly, in Maier's [26] two string problem, where the task is to tie two strings together that are hanging from the ceiling just out of arm's reach from each other using various objects available (e.g., a chair, a pair of pliers, etc.), people often fail to recognize the solution of tying the pair of pliers to one string and swinging it like a pendulum and catching it while standing on a chair between the strings. These findings demonstrate the phenomenon of "functional fixedness," where individuals have difficulty seeing unusual alternative uses for an artifact.

Another potentially relevant finding in the psychological literature is that individuals who acquire experience with classes of information and procedures tend to represent them in relatively large, holistic "chunks" in memory, organized by deep functional and relational principles [27–29]. Many researchers have argued that this ability to "chunk" underlies expertise and skill acquisition [27,30,31]. However, if the task at hand requires the individual to perceive or represent information in novel ways, e.g., to stimulate creative ideation in design, representation of that information in chunks might become a barrier to success, particularly if processing of component parts of the information chunks helps with re-representation [32–34].

These findings lead to a hypothesis that less-common example designs, which designers are less likely to have been exposed to, might present a unique advantage over more-common example designs in terms of the potential for stimulating creative ideation. Specifically, it could be that less-common examples are more likely to support multiple interpretations, and thus facilitate broader search through the space of possible solutions. Additionally, given that the commonness of example designs in the world (e.g., in practice, curriculum, etc.) is related to its representation in designers' long-term memory, e.g., ease/probability of recall, one could hypothesize that less-common examples might confer an advantage in terms of the novelty of solution paths they inspire. However, the literature gives no a priori reason to expect effects of commonness on mean quality of solution concepts.

2.3 Modality of Example Designs. With respect to the question of optimal reasoning modalities, a potential variable of interest is the contrast between pictorial and text-based representations of examples. One possible reason to investigate this contrast is that pictorial representations, e.g., sketches, photographs, and engineering drawings, often contain a higher degree of superficial features than text-based representations of the same information. This might be detrimental to conceptual design, as the presence of representations with a high degree of superficial detail, such as in detailed prototypes, in the physical design environment tend to restrict the retrieval of far-field analogies from memory [7]. On the other hand, some investigators argue that pictorial-based representations are better for conceptual design; for example, it has been shown that novice designers who are presented with sketches of example designs produce more novel and higher quality solution concepts on average relative to being presented with text-based example designs [35]. At a pragmatic level, too, in creating design-by-analogy tools, one ultimately has to decide on a representation format for potential analogies; thus, it is important to investigate if it matters whether they are represented in pictorial or text-based formats [10,11]. Additionally, it is important to know if the effects of example analogical distance or commonness are modulated by their representation modality.

2.4 Summary. In summary, a review of the relevant psychological literature suggests that investigating variations in example analogical distance, commonness, and modality might shed some important light on the questions regarding what to analogize over and whether there are optimal reasoning modalities. Prior work tentatively supports a hypothesis favoring far-field over near-field examples. With respect to commonness, to our knowledge, no studies have directly tested the effects of example commonness on conceptual ideation; however, the literature does suggest a hypothesis favoring less-common over more-common examples. Importantly, the theoretical and empirical literature suggest that there might be different effects of example analogical distance and commonness along different dimensions of the ideation process, thus motivating a fine-grained analytic approach to ensure that the effects of these variables can be clearly understood. Finally, the literature appears to be relatively equivocal about the contrast between pictorial and text-based representations; thus, our investigation of this variable in the present study is more exploratory than hypothesis-driven.

Table 1 Distribution of participants across conditions

	Near-field		Far-field	
	More-common	Less-common	More-common	Less-common
Picture	13	17	15	16
Text	17	16	16	17
Control		24		

3 Experimental Methods

3.1 Design. To investigate the effects of example analogical distance and commonness on conceptual design processes and possible interactions with modality, we conducted a 2 (distance: far-field vs. near-field) ×2 (commonness: more-common vs. less-common) ×2 (modality: pictures vs. text) factorial experiment, where participants, i.e., senior-level engineering students, were given a real-world design problem and were asked to generate solution concepts first briefly without examples, such that they understood the problem, and then with examples, to evaluate the effects of examples on problem solving. To establish whether examples of different types enabled or hindered problem solving, a control group of students executed a similar procedure but received no examples.

3.2 Participants. Participants were 153 students (predominantly mechanical engineering undergraduates) enrolled at two research universities in the United States. Participants were recruited from classes and were given either extra credit or compensation of $15 for their participation. Participants ranged from 20 to 38 years in age (M = 22, SD = 1.89). 70% were male. 87% were undergraduate engineering students (95% mechanical engineering, 5% electrical engineering and others) and 13% masters students in disciplines related to product design (e.g., mechanical engineering, product development, business administration). 66% of the participants had at least 1–6 months of engineering internship experience, and all but 2 out of the 153 students had experience with at least one prior design project in their engineering curriculum. Approximately 82% of the students had taken at least one course where a structured approach to design was taught. Thus, most of the participants had relevant mechanical engineering domain knowledge and design experience.

Participants were randomly assigned to one of the nine possible conditions in each class by distributing folders of paper materials prior to students arriving in class. The obtained distribution of participants across the nine conditions is shown in Table 1—the sample populations, *N*s, are unequal not because of dropout but rather from stochasticity in where students chose to sit down. With these sample populations, statistical power for detecting three-way interactions (not our theoretical goal) is modest, but power for detecting two-way interactions and main effects is good.

3.3 Design Problem. The design problem was to design a low cost, easy to manufacture, and portable device to collect energy from human motion for use in developing and impoverished rural communities, e.g., India, many African countries. This design problem was selected to be meaningful and challenging to our participants. The problem was meaningful in the sense that real-world engineering firms are seeking solutions to this problem and the problem involves social value; thus, students would be appropriately engaged during the task [36–38]. The problem was challenging in the sense that a dominant or accepted set of solutions to the problem has yet to be developed (so students would not simply retrieve past solutions), but it was not so complex as to be a hopeless task requiring a large design team and very detailed task analysis.

3.4 Selection of Examples. Examples were patents selected from the U.S. Patent Database. Candidate patents were retrieved using keyword search on the U.S. Patent and Trade Office website. The keywords used were basic physical principles, such as *induction*, *heat transfer*, *potential energy*, as well as larger categorical terms like *mechanical energy*. The final set of eight patents was selected by two PhD-level mechanical engineering faculty based on two sets of criteria: (1) balanced crossing of the analogical distance and commonness factors, such that there would be two patents in each of the four possible combinations, and (2) overall applicability to the design problem, over and above analogical distance and commonness. Each participant in the analogy conditions received two examples of a particular type, roughly balanced across conditions for applicability. The patents for each of the conditions are shown in Table 2.

With respect to the first set of criteria, the specific guidelines for selection were as follows:

1. *Distance*: Far-field patents were devices judged to be not directly for the purpose of generating electricity, while near-field patents were those judged to be directly for the purpose of generating electricity.
2. *Commonness*: More-common patents were devices judged likely to be encountered by our target population in their standard engineering curriculum and/or everyday life, while less-common patents were those judged unlikely to be seen previously by the participants under typical circumstances.

With respect to the modality factor, in the picture conditions, participants received a representative first figure from the patent, which typically provides a good overview of the device, while in the text conditions, participants received the patent abstract. In some cases, abstracts differed substantially in length; to equate for quantity of text across conditions, overly brief abstracts were augmented with additional text from the body of the patent, which elaborated on the details of the design and technology. To provide some foundational context, all text-and-picture-condition participants also received the patent title.

3.5 Experimental Procedure. The experiments were conducted during class. Participants generated solution concepts in three phases and subsequently completed a background survey. Participants proceeded through the phases using a sequence of envelopes to carefully control timing of the task and exposure to examples across conditions. In particular, we wanted to ensure that design examples were received only after participants had made some substantial progress in ideation, since prior work has shown that examples and potential analogies are most helpful when received after ideation has already begun [39,40]. The overall time allowed for this task was sufficient to allow for broad exploration of the concept space, but not enough to develop

Table 2 Patents for each condition

	Near-field	Far-field
More-common	-Waterwheel-driven generating assembly (6208037) -Recovery of geothermal energy (4030549)	-Escapement mechanism for pendulum clocks (4139981) -Induction loop vehicle detector (4568937)
Less-common	-Apparatus for producing electrical energy from ocean waves (4266143) -Freeway power generator (4247785)	-Accelerometer (4335611) -Earthquake isolation floor (4402483)

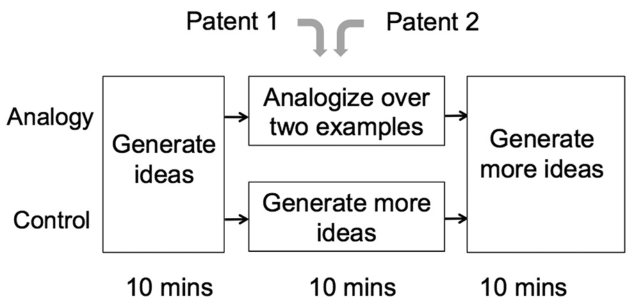

Fig. 1 Comparison of experimental procedures for analogy vs. control groups

particular ideas in depth, matching our focus on the ideation process.

Analogy and control groups executed the same overall sequence, but differed in the particular activities in the second phase of ideation (see Fig. 1 for a comparison of the procedures). In general, the sequence of phases was to: (1) read design problem and generate solution concepts, (2) either (a) review two patents and write/draw solutions/ideas that come to mind when looking at the patents or (b) continue generating concepts, and (3) generate more solution concepts. Each phase lasted 10 min.

With respect to idea generation, participants were instructed to generate and record as many solution concepts to the design problem as they could, including novel and experimental ones, using words and/or sketches to describe their solution concepts.

4 Ideation Metrics

The experiment generated 1321 total ideas. To thoroughly explore the range of effects of varying the analogical distance, commonness, and modality of design examples on conceptual design processes, we applied a range of ideation metrics to these ideas: (1) the extent to which solution features were transferred from examples, (2) quantity of ideation, (3) breadth of search through the space of possible solutions, (4) quality of solution concepts, and (5) novelty of solution concepts. The first three metrics provided measures of the ideation process of participants and how they processed the examples: examining *solution transfer* provides insight into the mechanisms by which participants might be stimulated by the examples, e.g., did they actually use solution elements; measuring *quantity* of ideation gave a sense of how participants were exploring the design space, i.e., whether they were generating and refining a small number of ideas, or exploring multiple concepts and variations of concepts, which is associated with higher likelihood of generating high-quality concepts [4]; finally, *breadth of search* was taken to be a measure of the ability to generate a wide variety of ideas, which is associated with the ability to restructure problems, an important component of creative ability [41–43]. The final two metrics focused on the ideation products of participants. We investigated quality because in design, a baseline requirement is that concepts must meet customer specifications; design concepts that are novel but do not meet customer specifications cannot be considered acceptable designs, let alone creative ones [41]. We investigated novelty because there is a high degree of consensus in the literature that creative products are at least novel [41,42].

4.1 Data Preprocessing. The raw output of each participant was in the form of sketches and/or verbal descriptions of concepts. Examples of participant-generated solution concepts are shown in Fig. 2. A number of preprocessing steps were necessary to prepare the data for coding and analysis.

First, each participant's raw output was segmented by a trained coder into solution concepts. A sketch and/or verbal description was segmented as one solution concept if it was judged to describe one distinct solution to the design problem. Variations of solutions (e.g., with minor modifications) were counted as distinct solution concepts. Segmentation was independently checked by a second coder. Inter-rater agreement was high (96%), and all disagreements were resolved by discussion. Next, sets of two senior mechanical engineering students rated each solution concept as meeting or not meeting the minimum constraints of the design problem, as described above, to remove off-topic inspirations generated by the patent examples, especially in the second phase. Inter-rater agreement was acceptable, with an average Cohen's kappa of 0.72. All disagreements were resolved through discussion. The 1066 solution concepts remaining after preprocessing constituted the final data set for analysis.

4.2 Solution Transfer. Solution transfer was defined as the degree to which a given participant's idea set contained solution features from the examples she/he received. The process of producing a solution transfer score for each participant was as follows. First, key features were generated by one of the co-authors for each of the eight patent examples, and the list was cross-checked for relevance by the other co-authors. Recall that each participant received two examples; however, since picture and text examples were essentially the same examples (only in different representations), the $2 \times 2 \times 2$ design reduced to a 2×2 design, leaving a total of eight examples. A total of 39 key features were identified. Because some features overlapped across examples (e.g., "built into ground, stationary, or permanent" was

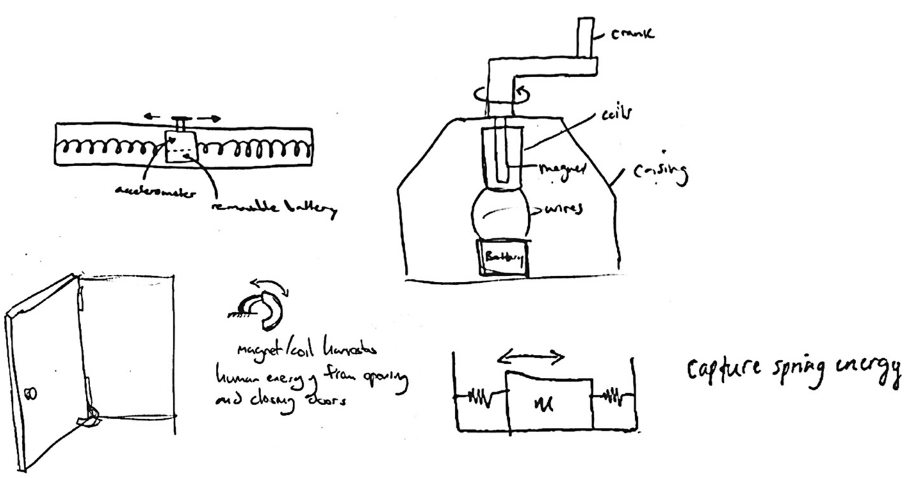

Fig. 2 Example participant solution concepts

associated with four patent examples), there was not a simple one-to-one mapping of features to examples. The number of features associated with each of the eight examples ranged from 4 to 7 ($M = 4.9$, $SD = 1.0$). Second, each participant solution concept was coded for the presence or absence of a set of the features found in the full set of patent examples presented to participants. The first 50% of solution concepts was double-coded by two senior mechanical engineering students to establish reliability. Later, all coding was completed by one student only. Test-retest measures of reliability were obtained in lieu of inter-rater reliabilities. Cohen's kappa averaged across features was 0.57. Because some features had low coding reliability or high overlap of features across many of the patents or simply were common elements of most proposed solutions across all conditions, the initial set of 39 features was filtered down to 23 features according to three criteria:

1. Acceptable inter-rater agreement, i.e., Cohen's kappa greater than 0.4.
2. Not shared by more than three examples.
3. Not too common, i.e., base rate (collapsed across conditions) less than 0.5.

After filtering, the number of features ranged from 1 to 5 ($M = 2.9$, $SD = 1.4$) per example and from 4 to 8 ($M = 5.8$, $SD = 1.7$) per each of the four conditions in the distance by commonness 2×2 design. Cohen's kappa averaged across the filtered set of features was 0.66.

To produce solution transfer scores for each participant, the following procedure was used. First, for each cell in the 2×2 (distance $\times$ commonness), we computed for each participant the proportion of his/her ideas that had at least one solution feature from the examples she/he received. Next, this proportion was converted into a standardized z-score by subtracting the mean and dividing by the standard deviation of proportion scores for all participants who were *not* in that 2×2 cell. The reason for using this transformation was that solution features from examples could occur in participants' ideas even if they never saw the relevant examples; this transformation allows us to separate the probability of participants using solution features from examples they have seen from the probability of using those solution features even if they had never seen the examples. For each participant, the transfer score was the z-score of each feature relevant to the examples they actually received.

The solution transfer score thus gave a measure of the degree to which a given participant's idea set differed from "normal" in terms of the proportion of ideas with at least one feature from the examples she/he received. To illustrate, suppose participant 1001 had a z-score of 1.34 for far-field, more-common examples. This number would say that the proportion of 1001's ideas with at least one solution feature from the examples s/he received was 1.34 standard deviations higher than the mean proportion of ideas with at least one solution from those examples under "normal" circumstances (i.e., without having seen either of the two far-field, more-common examples).

4.3 Quantity of Ideation. Quantity of ideation was defined as the number of solution concepts generated post analogy, i.e., from the second phase of ideation onwards, that met the minimum constraints of the design problem, viz. (1) the device generates electricity, and (2) it uses human motion as the primary input. As noted in the introduction, quantity is often taken to be a key component of creativity. Quantity was defined at the level of the participant, i.e., each participant received a single quantity score. Because we were primarily interested in the effects of examples on quantity, analyses concentrated on the number of solution concepts generated after receiving examples (i.e., after the first phase) adjusting for the number of solution concepts generated in the first phase (which acted as a covariate to adjust for baseline variation in quantity across participants).

4.4 Breadth of Search. Breadth of search was conceptualized in our study as the proportion of the space of possible solutions searched by a given participant. To determine the space of possible solutions, the design problem was first functionally decomposed into potential subfunctions by one of the authors, drawing from the reconciled function and flow basis of Hirtz and colleagues [16].

Due to the open-ended nature of the design problem, a relatively large number of subfunctions were initially generated, as follows:

1. Import/accept human interaction
2. Transform human energy to mechanical energy
3. Transform human energy to alternative energy
4. Import other material
5. Contain/store other material
6. Transfer other material
7. Import alternative energy source
8. Transform alternative energy source into mechanical energy
9. Transform alternative energy source to alternative energy
10. Transform collected energy to mechanical energy
11. Transmit mechanical energy
12. Transform mechanical energy
13. Store mechanical energy
14. Transform mechanical to alternative energy
15. Transform alternative energy to electrical energy
16. Actuate/deactuate energy
17. Transform mechanical energy to electrical energy
18. Condition electrical energy
19. Store electrical energy
20. Supply electrical energy
21. Transmit electrical energy
22. Convert electrical to light or EM

Each subfunction solution consisted of a *how* and *what* component, where the former specifies the component of the solution concept that implements the subfunction, and the latter specifies either the input or the output of the subfunction (whichever is the less specified). For example, a solution for the subfunction "import human" might be "*foot* with *pedals*."

Two senior mechanical engineering students independently coded the solutions to the subfunctions for each solution concept. The solution types for the *how* and *what* components of each subfunction were generated bottom-up by the students as they coded, with each new solution type being added to a running list of solution types; the running list of solution types for each subfunction constituted the coding scheme. Inter-rater reliability was high, with an average Cohen's kappa across subfunctions of 0.84. All disagreements were resolved by discussion.

While the nature of the design problem was open-ended, a core set of subfunctions emerged from the dataset: only a small subset of the initial set of subfunctions occurred often enough for stable estimates of breadth and novelty (i.e., base rate greater than 0.1, collapsed across conditions):

1. Import human
2. Transform human energy to mechanical energy
3. Import alternative energy
4. Transform alternative energy to mechanical energy
5. Transform mechanical energy to electrical energy
6. Store electrical energy

Upon more detailed analysis, it turned out that there were only two solution types for the subfunction "store electrical energy," namely "battery" or "capacitor," and the frequency of occurrence for each solution type was relatively equivalent; thus, novelty scores for this subfunction would be unlikely to differentiate between participants. Furthermore, since the design problem was focused on the problem of harvesting (vs. storing) energy, data for this subfunction were not included in computations of breadth.

We defined the space of possible solutions for each of the *what* and *how* components of each subfunction by enumerating the number of distinct solution types generated by participants across all phases of ideation. A breadth score b_j for each participant on subfunction j was then computed with

$$b_j = \sum_{k=1}^{n} w_{jk} \times \frac{C_{jk}}{T_{jk}} \quad (1)$$

where C_{jk} is the total number of solution *types* generated by the participant for level k of subfunction j, T_{jk} is the total number of solution *types* produced by *all* participants for level k of subfunction j, and w_k is the weight assigned level k. To give priority to breadth of search in the *what* space (types of energy/material manipulated), we gave a weight of 0.66 to the *what* level (which was assigned to $k=1$), and a weight of .33 to the *how* level (which was assigned to $k=2$). An overall breadth score for each participant was given by the average of breadth scores for each of the three subfunctions j.

4.5 Quality. Quality of solution concepts was measured using holistic ratings on a set of subdimensions of quality. Two other senior mechanical engineering students independently coded solution concepts on 5-point scales ranging from 0 to 4 (0 is unacceptable and 4 is excellent) for six subdimensions of quality, corresponding to a set of possible customer specifications:

1. Cost
2. Feasibility of materials/cost/manufacturing
3. Feasibility of energy input/output ratio
4. Number of people required to operate device at a given moment
5. Estimated energy output
6. Portability
7. Time to set up and build, assuming all parts already available at hand

These subdimensions were generated by the second author, who is a Ph.D. candidate in mechanical engineering focusing on design methods and cognition, and checked for validity by two other authors, who are mechanical engineering faculty specializing in engineering design. For each subdimension, each point on the 5-point scale was anchored with a unique descriptor. For example, for the "feasibility of energy input/output ratio" subdimension, 0 was "unfeasible design or input energy completely dwarfs output," 1 was "input less than output", 2 was "I/O about even," 3 was "sustainable/little surplus output; human input easy," and 4 was "output significantly higher than input." Inter-rater agreement was computed using a Pearson correlation between the ratings of the two coders for each subdimension. The average of correlations across subdimensions was 0.65, and the range was from 0.49 to 0.77. An overall quality score was computed for each solution concept, as given by

$$Q = \frac{\sum_{j=1}^{n} q_j \times r_j}{Q_{\max}} \quad (2)$$

where q_j is the quality score for quality subdimension j, r_j is the reliability of the coding for that subdimension, and Q_{max} is the maximum possible overall quality score, which would be given by setting q_j to 4 for each subdimension. The contributions of subdimension scores to the overall quality score were weighted by reliability to minimize the influence of measurement error. Since the overall quality score was a proportion of the maximum possible quality score, the score ranged from 0 to 1. Agreement between coders at the level of this composite score was acceptable ($r=0.68$).

4.6 Novelty. Novelty was defined as the degree to which a particular solution type was unusual within a space of possible solutions. This approach allowed us to avoid the difficulties of judging the novelty of thousands of solution concepts via holistic rating methods. Recall that for the breadth metric, the space of possible solutions was defined in terms of a set of five core subfunctions for the design problem; recall further that each subfunction was decomposed further into *what* and *how* components, where the former specifies the component of the solution concept that implements the subfunction, and the latter specifies either the input or the output of the subfunction (whichever is the less specified). Rather than computing novelty scores for solutions to each level of each subfunction (the *what* and *how* levels), we chose to compute novelty scores for the *conjunction* of *what* and *how* solution components for each subfunction. For example, rather than computing the relative unusualness of the solution components "foot" and "pedals" separately for the solution "foot with pedals" for the subfunction "import human interaction," the relative unusualness of the solution "foot with pedals" relative to other solutions would be computed. The rationale for this choice was that these words in conjunction as a solution have a specific meaning that needed to be considered. Novelty scores were computed for each subfunction solution using Eq. (3), which is a formula adapted from Ref. [39]

$$N_i = \frac{T_i - C_i}{T_i} \quad (3)$$

where T_i is the total number of solution *tokens* generated for subfunction i in the first phase of ideation (collapsed across all participants), and C_i is the total number of solution tokens of the current solution *type* in the first phase of ideation. Because this measure was essentially a measure of proportion, the novelty score for each idea ranged from 0 to 1, with 0 representing solution types found in every solution (this extreme was never observed) and 1 representing solution types that never occurred in the first phase. The initial set of solution concepts (generated in the first phase of ideation) was taken to be the original design space of the participants since it corresponded to concepts generated prior to receiving examples. The final novelty score for each solution concept was the average of its subfunction novelty scores.

5 Results

5.1 Relationships Between Metrics. Analysis of the interrelationships between the ideation metrics suggested a preliminary process model that could account for these correlations and help to conceptually organize the results (see Fig. 3). Of course, correlations per se do not guarantee causation and other causal models are possible.

The preliminary process model is as follows:

- Increased solution transfer results in decreased quantity, possibly because many participants had trouble thinking of solutions beyond the ones presented.
- A high quantity of ideation allows for greater breadth of search, even if only on a statistical sampling basis.

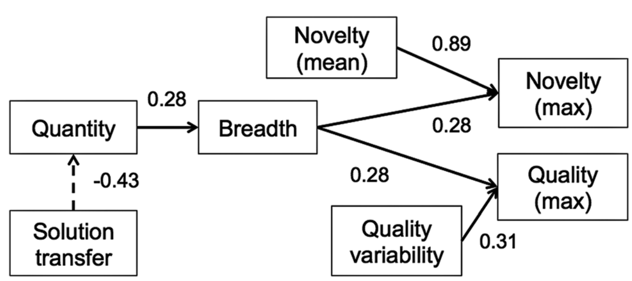

Fig. 3 Summary of intermetric correlations. Numbers shown are Pearson's *r*. All correlations are significant at $p<0.01$.

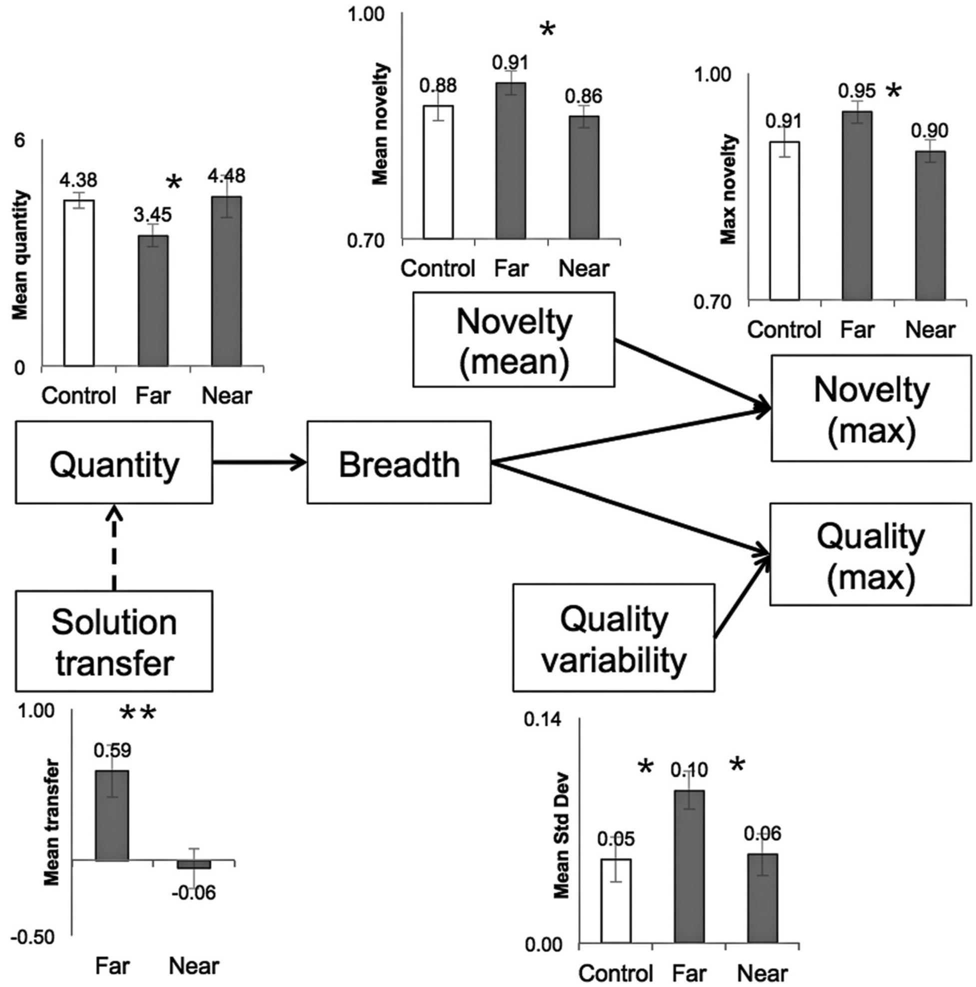

Fig. 4 Summary of effects of example distance. *$p<0.05$ and **$p<0.01$. Control group data are shown in white bars. Error bars are ±1 standard error.

- Greater breadth of search, perhaps also only on a statistical sampling basis, in turn allows for the generation of higher novelty and higher quality solution concepts.
- Repeatedly searching on the fringes of the design space (as measured by high average novelty) further increases the probability of finding a highly novel concept.
- Finally, increasing the variability of the quality of solution concepts increases the probability of generating a high-quality solution concept. This last relationship is in accord with the work of Ulrich and colleagues in the field of innovation management, who have argued and showed empirically that one way to increase the likelihood of finding high market potential product concepts is to increase the variance of the quality of the concepts that are generated [4,44].

5.2 Effects of Analogy Manipulations on Ideation Metrics. We now present our findings by manipulation (distance, commonness, and modality), using the preliminary process model as an organizational framework. Effects of manipulations on the ideation metrics will be described following the flow of the process model, first considering solution transfer, quantity, and breadth, followed by consideration of effects on quality and novelty of ideation. Separate 3-way (distance × commonness × modality) analysis of variance (ANOVA) models were computed for each process variable in the model. In some cases (indicated in each case), the level of that variable during the pre-analogy phase was used as a covariate in the analysis because the baseline measure was a significant predictor of postanalogy performance.

5.2.1 Analogical Distance of Examples. There was a main effect of example distance ($p<0.01$, $\eta^2=0.08$) on solution transfer, where participants who received far-field examples were much more likely than participants who received near-field examples to use solution elements from the examples they received ($d=0.60$);[1] in fact, solution features from near-field examples were no more likely to be present in participant solutions after processing examples relative to the pre-example phase (see Fig. 4, bottom left).

There was also a main effect on quantity ($p<0.01$, $\eta=0.05$), where participants who received far-field examples generated significantly fewer solution concepts relative to participants who received near-field examples ($p<0.05$, $d=-0.30$; see Fig. 4, upper left). There were no significant differences in terms of quantity between receiving no examples (control) and receiving either far- or near-field examples. However, the small effect of distance on quantity did not translate into an effect on breadth: there were no reliable effects of distance on breadth of search ($p=0.78$, $\eta^2=0.00$).

With respect to quality of solution concepts, there were no effects of distance on either mean or maximum quality. However, there was a main effect of distance of the variability in quality of participants' solution concepts ($p<0.05$, $\eta^2=0.06$; see Fig. 4, lower right), where participants who received far-field examples had a larger standard deviation in quality of solution concepts

[1] d statistics estimate the size of the difference in group means in terms of the average standard deviation of the two groups in the contrast; in this case, $d=0.60$ estimates that the mean probability of transfer is greater with far-field vs nearfield examples by 0.60 of a standard deviation (a moderate to large difference).

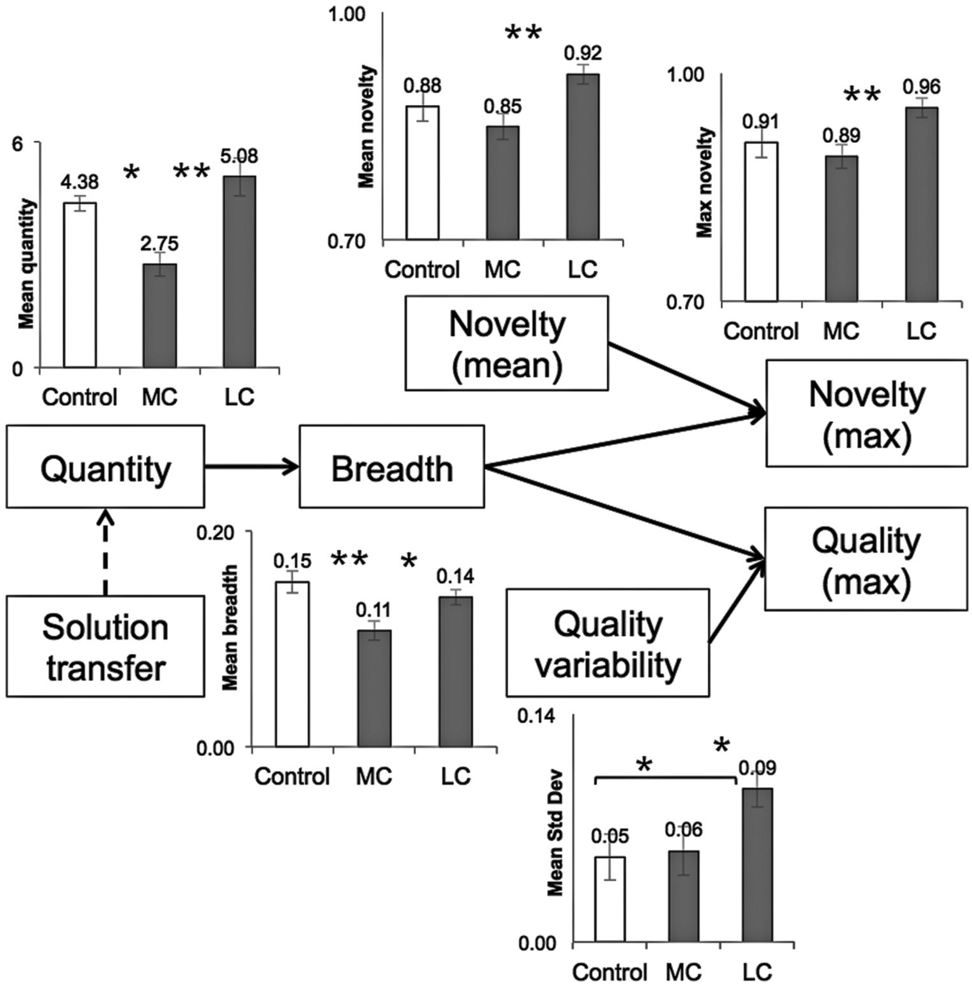

Fig. 5 Summary of effects of example commonness. *$p<0.05$and **$p<0.01$. Control group data are shown in white bars. Error bars are ±1 standard error.

than participants who received either near-field examples ($p<0.05$, $d=0.64$) or no examples ($p<0.05$, $d=0.78$). There were no significant differences between receiving near-field examples vs. no examples.

Finally, there was a main effect of distance on mean novelty ($p<0.05$, $\eta^2=0.04$), where participants who received far-field examples generated solution concepts that were more novel on average relative to participants who received near-field examples ($p<0.05$, $d=0.56$; see Fig. 4, upper right). Similar patterns of effects were found with maximum novelty of solution concepts ($p<0.05$, $\eta^2=0.04$), where the most novel solution concept of participants who received far-field examples was more novel on average relative to the most novel solution concept of participants who received near-field examples ($p<0.05$, $d=0.56$). There were no significant differences between participants who received no examples (control) vs. near- or far-field examples on either mean or maximum novelty.

In summary (see Fig. 4), example distance appeared to have significant effects on multiple aspects of ideation. Specifically, novelty and variability in quality of concepts increased as a function of receiving far-field examples, although only in the latter case was the contrast with control statistically significant. The solution transfer metric suggests that these increases might be associated with incorporating solution elements from the far-field examples. However, the benefits of far-field examples came with a slight cost, viz. a reduction in quantity: in meaningful terms, the cost of processing far-field examples given a standard time for ideation appeared to be, on average, about one solution concept.

5.2.2 Commonness of Examples. Turning now to the main effects of commonness in the same ANOVAs, there were no reliable effects on *solution transfer* (p = 0.30, $\eta^2=0.01$). However, there was a main effect on *quantity* ($p<0.01$, $\eta^2=0.12$), where participants who received more-common examples generated significantly fewer solution concepts relative to participants who received either more-common examples ($p<0.01$, $d=-0.67$) or no examples ($p<0.01$, $d=-0.76$; Fig. 5, upper left). There were no significant differences in quantity between participants who received less-common vs. no examples (control). There was also a main effect of on *breadth of search* ($p<0.01$, $\eta^2=0.07$), where participants who received more-common examples searched less of the design space than participants who received either less-common examples ($p<0.05$, $d=-0.61$; Fig. 5, lower middle) or no examples ($p<0.01$, $d=-1.03$). There were no significant differences in breadth of search between participants who received less-common vs. no examples (control).

With respect to *quality* of solution concepts, there were no reliable effects of commonness on either mean or max quality. However, there was a main effect on variability in quality of participants' solution concepts ($p<0.05$, $\eta^2=0.06$; see Fig. 5, lower right), where participants who received less-common examples had a larger standard deviation in quality of solution concepts than participants who received either more-common examples ($p<0.05$, $d=0.62$) or no examples ($p<0.05$, $d=0.68$). There were no significant differences between receiving more-common examples vs. no examples.

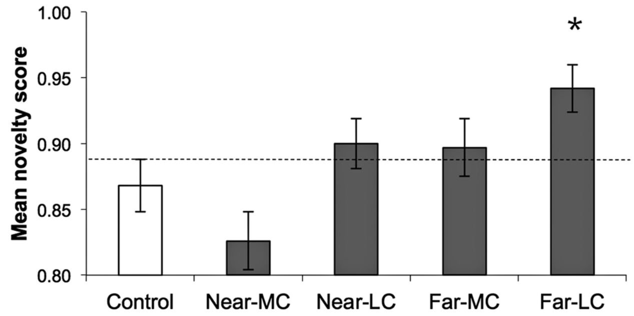

Fig. 6 Mean novelty of solution concepts by example distance and commonness. *$p<0.05$. Error bars are ±1 standard error.

Finally, there were main effects on mean *novelty* ($p<0.01$, $\eta^2=0.10$), where participants who received less-common examples generated solution concepts that were more novel on average relative to participants who received more-common examples ($p<0.01$, $d=0.61$; see Fig. 5 upper right) and maximum novelty ($p<0.01$, $\eta^2=0.96$), where the most novel solution concept of participants who received less-common examples was more novel on average relative to the most novel solution concept of participants who received more-common examples ($p<0.01$, $d=0.61$). There were no significant differences between participants who received no examples (control) vs. more- or less-common examples on either mean or maximum novelty.

In summary (see Fig. 5), example commonness also appeared to have significant effects on ideation. Less-common examples were associated with more positive ideation processes and products relative to more-common examples, with benefits for quantity and breadth of ideation, variability in solution quality, and novelty of solution concepts, although only in the case of variability in solution quality was the contrast with control statistically significant.

5.2.3 Joint Effects of Example Distance and Commonness on Novelty. While far-field and less-common examples separately increased novelty of ideas, neither far-field examples as a whole nor less-common examples as a whole were significantly different from control, which sat in the middle. To examine whether the *combination* of far-field and less-common properties increased novelty over control, we used a Dunnett's multiple comparison post hoc test. Since there were no effects of modality on novelty (described below), we collapsed across the picture and text factors and conducted the post hoc test comparing each of the combinations in the 2×2 matrix (distance x commonness) with the control condition as a reference group. The post hoc test showed that the combination of far-field, less-common examples did in fact increase novelty vs. control, for both mean ($d=1.14$; see Fig. 6) and max ($d=1.29$).

5.3 Effects of Example Modality. Turning to the effects of modality in the overall ANOVAs, there was a main effect of example modality ($p<0.01$, $\eta^2=0.09$) on solution transfer, where participants who received their examples in text form were more likely to use solution elements from the examples they received, regardless of distance or commonness of the example ($d=0.60$; Fig. 7, lower left).

There was also a main effect of on quantity ($p<0.01$, $\eta=0.12$; Fig. 7, upper left), where participants who received text examples generated significantly fewer solution concepts relative to participants who received either picture examples ($p<0.01$, $d=-0.67$) or no examples (control; $p<0.05$, $d=-0.56$). There were no significant differences between participants who received picture examples vs. no examples (control). Thus, receiving examples in text form increased the likelihood of being able to use solution

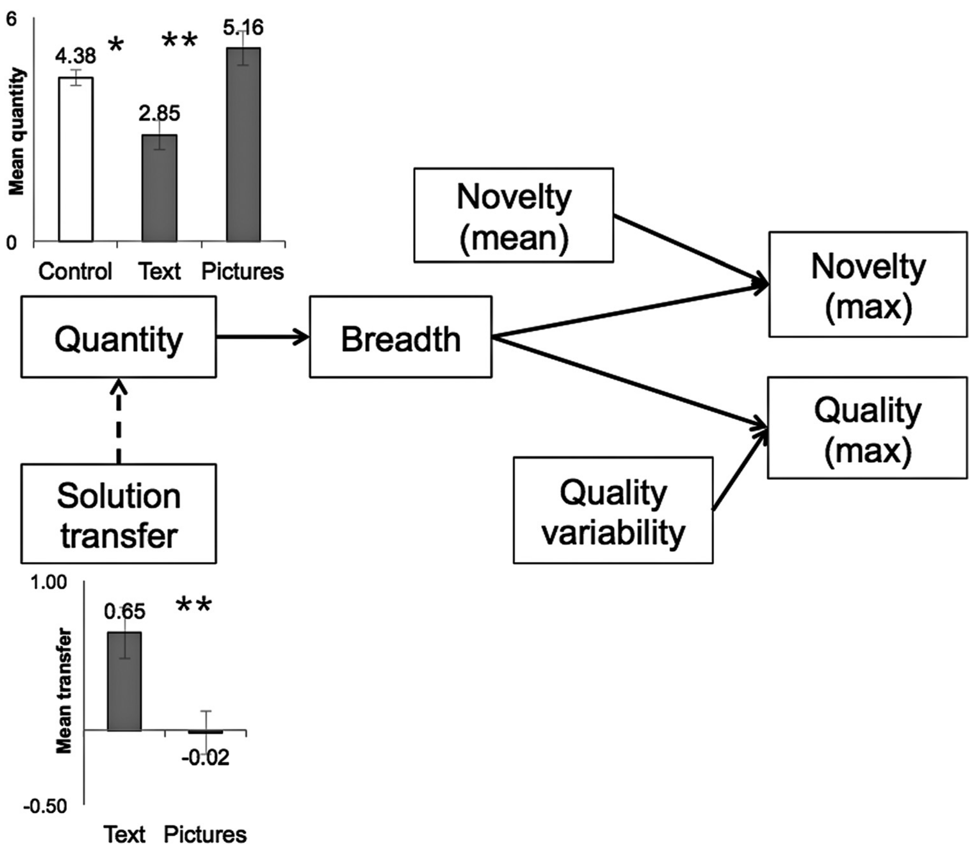

Fig. 7 Summary of effects of example modality. *$p<0.05$ and **$p<0.01$. Control group data are shown in white bars. Error bars are ±1 standard error.

elements from those examples relative to picture form, but also decreased quantity by an average of about two concepts relative to receiving either picture or text examples.

There were no additional effects of modality on the other dependent measures (breadth, $p = 0.11$, $\eta^2 = 0.03$; mean novelty, $p = 0.20$, $\eta^2 = 0.02$; max novelty, $p = 0.49$, $\eta^2 = 0.00$; quality variability, $p = 0.44$, $\eta^2 = 0.01$). Thus, modality had little impact on the key end-state outputs of the ideation process, unlike the effects of example commonness or example analogical distance.

6 Discussion

6.1 Optimal Example Types. Our findings demonstrate that the analogical distance and commonness of examples significantly influences their impact on designers' ideation. With respect to analogical distance, augmenting ideation with far-field examples brings significant benefits vis-à-vis the kinds of concepts that can be generated; specifically, ideation with far-field examples enhances the ability to generate highly novel solution concepts and also allows for more variability in the quality of concepts, which may increase the likelihood of generating high quality concepts. It is interesting to note that, even though the far-field examples we gave participants were not energy-generating devices, they were still able to benefit from the concepts and solution elements in the devices. This sort of transfer is greater in distance than typically seen in the analogy literature, where far-field analogies in problem solving are usually from cases in other domains that are surface dissimilar but still solve the same basic problem [20,45].

However, the use of far-field examples was not without some cost. Far-field examples reduced overall quantity of ideation relative to near-field or no examples. This finding can be interpreted in terms of processing difficulty. When we computed an additional 3-way ANOVA model on quantity for only the final phase of ideation, removing from consideration quantity of ideation while processing examples, the effects of distance were no longer present ($p = 0.47$). This suggests that the reduction in quantity can be attributed to the time taken to map the far-field example to the design problem. Thus, it appears that far-field examples not only carry with them the potential to increase novelty and quality of design concepts generated but also carry an initial processing cost in terms of time taken to map them to the target problem.

With respect to commonness of examples, we found that the use of less-common examples positively impacts ideation. Less-common examples resulted in increased quantity of ideation, breadth of search, and higher novelty of ideas relative to more-common examples. In a follow-up analysis analyzing quantity for only the final phase of ideation, the positive effects of less-common examples relative to more-common examples were still present ($p < 0.05$, $d = 0.56$), suggesting that the effects cannot be explained simply in terms of initial processing costs, as in the case of distance effects on quantity. Thus, it seems that less-common examples might be more beneficial for stimulating ideation, particularly in terms of novelty of concepts generated. This finding is in accord with some work in the domain of artistic creativity, where it has been shown that copying novel artworks has a positive effect on the ability of art students to flexibly re-interpret artwork and increases the novelty of the artworks produced [46].

While distance and commonness had some similar effects on ideation processes and products, our fine-grained analytic approach suggests some potentially important distinctions. The critical contrast seems to be with respect to effects on quantity and breadth of ideation. Far-field examples increased novelty of solutions and variability in solution quality, but appeared to do so via solution transfer, and resulted in decreased quantity; in contrast, less-common examples also increased novelty and quality variability, but appeared to do so via broadening the search space and increasing quantity. One way to interpret this contrast is that example distance and commonness have different mechanisms of inspiration. Based on the results, one could hypothesize that far-field examples inspire designers by moving them into one or two novel regions of the design space (high solution transfer, high novelty), which they then explore in more depth (low quantity, no benefits on breadth); in contrast, one could hypothesize that less-common examples inspire designers by moving them into multiple different regions of the design space via re-interpretation of design functions and features (low solution transfer, high breadth, and quantity).

6.2 Optimal Representation Modality of Examples. With regard to the outcome measures of novelty and quality of solution concepts, we found that the representation modality of examples did not change the effects of the distance and commonness factors on ideation. However, we did find evidence for a negative effect of text representations on overall quantity of ideation relative to picture or no examples. Similar to the effects of distance on quantity, this suppression effect of text representations can be interpreted in terms of initial processing costs: when we analyzed only the last phase of ideation, the effect of modality was weaker (pictures vs. text, $d = 0.32$; pictures vs. control, $d = 0.45$) and no longer statistically significant ($p = 0.07$). As an ancient proverb puts it, one picture may be worth 10,000 words with respect to conveying design concepts.

6.3 Caveats. The current work comes with a number of caveats. First, we have examined only one design problem. Although a real design problem of some complexity, examples may have different effects on more complex design problems. Second, we examined the effects of particular examples rather than a range of examples sampled multiple times from a class of examples. This experimental design choice made it more feasible to analyze solution transfer but raises possibilities of effects being caused by odd examples or example descriptions. To reduce this threat, we had two examples per condition, and the factorial design of the study permits for multiple replications of main effects. Third, our participants were senior-level engineering students, for the most part, rather than expert designers, and there is some research to suggest that novices have more difficulty with analogical mappings [5,47]. However, design teams sometimes include less experienced designers. Finally, our study focused only on the earliest ideation phase, and future work will have to examine the effects of examples on downstream, and in particular finished, solutions. This restriction was most salient in the analyses of quality in that many of the ideas were not feasible or not fleshed out sufficiently to determine feasibility. However, a number of studies point to early ideation as a key moment for intervention to generate innovative designs [3,4].

6.4 Practical Implications and Future Work. The overall focus of this study was on whether particular kinds of examples are more helpful than others for stimulating ideation. However, with the inclusion of a control group, which received no examples, we were able to answer a separate but related question: all things considered, does analogizing over examples confer benefits over and above ideating without examples? In other words, is design-by-analogy worth the extra time and effort? Our findings suggest that if the goal of conceptual ideation is to ultimately generate and develop a concept that is high quality *and* novel, then the answer is yes.

There are also implications for the design of tools and methods to support design-by-analogy. As noted in the introduction, a range of previous design-by-analogy methods have been developed; of particular interest is the development of computational tools that automate the search for analogies [48]. It is well known in the psychological literature that retrieving far-field analogies is cognitively difficult; remindings tend to be significantly constrained by surface similarity [49], reducing the probability of retrieving potentially relevant surface dissimilar analogies. Thus, computational tools that are able to define and compute functional and surface similarity between items in a design space in a principled manner relative to the current design problem would hold excellent potential as aids for inspiration. These tools might be able to maximize the potential benefits of analogies by retrieving and delivering to the designer in a timely manner surface dissimilar analogies and potentially (as our findings suggest) even analogies that do not necessarily provide direct solutions to the target problem. Additionally, if these systems are able to give priority to analogies that are relatively unusual or infrequently encountered, the potential for inspiration might be even higher.

Currently, the state of the art for computational design-by-analogy tools has not reached the point of being able to provide flexible and real-time support in this manner. The present work provides an impetus for investment into this important research area, as the potential benefits to engineering practice and to society via increased innovation is high

Acknowledgment

This work is supported by grants from the National Science Foundation, Grant Nos. CMMI-0855326, CMMI-0855510, and CMMI-0855293, and, in part, by the University of Texas at Austin Cockrell School of Engineering and the Cullen Trust Endowed Professorship in Engineering No. 1. Any opinions, findings, or recommendations are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] Pisano, G., and Shih, W., 2009, "Restoring American Competitiveness," Harv. Bus. Rev., **87**, pp. 114–125.
[2] National Academy of Engineering, 2005, *Engineering Research and America's, Future: Meeting the Challenges of a Global Economy*, National Academies Press, Washington, DC.
[3] Vogel, C. M., Cagan, J., and Boatwright, P., 2005, *The Design of Things to Come*, Wharton School Pub, Upper Saddle River, NJ.
[4] Terwiesch, C., and Ulrich, K. T., 2009, *Innovation Tournaments*, Harvard Business School Pub.
[5] Casakin, H., and Goldschmidt, G., 1999, "Expertise and the use of visual analogy: Implications for design education," Des. Stud., **20**(2), pp. 153–175.
[6] Goel, A., 1997, "Design, Analogy and Creativity," IEEE Expert, **12**(3), pp. 62–70.
[7] Christensen, B. T., and Schunn, C. D., 2007, "The Relationship of Analogical Distance to Analogical Function and Pre-Inventive Structure: The Case of Engineering Design. Mem. Cognit., **35**(1), pp. 29–38.
[8] Linsey, J., Murphy, J., Laux, J., Markman, A., and Wood, K. L., 2009, "Supporting Innovation by Promoting Analogical Reasoning," *Tools of Innovation*, Oxford University Press, New York, NY.
[9] Gentner, D., 1983, "Structure-Mapping: A Theoretical Framework for Analogy," Cogn. Sci., **7**, pp. 155–170.
[10] Linsey, J., Murphy, J., Markman, A., Wood, K. L., and Kortoglu, T., 2006, "Representing Analogies: Increasing the Probability of Innovation," *ASME International Design Theory and Method Conference*, Philadelphia, PA.
[11] Linsey, J., Wood, K. L., and Markman, A., 2008, "Modality and Representation in Analogy," Artif. Intell. Eng. Des. Anal. Manuf. (Special Issue on Multimodal Design), **22**(2), pp. 85–100.
[12] Gordon, W. J. J., 1961, *Synectics: The Development of Creative Capacity*, Harper and Brothers, New York.
[13] French, M., 1988, *Invention and Evolution: Design in Nature and Engineering*, Cambridge University Press, Cambridge, UK.
[14] Hacco, E., and Shu, L. H., 2002, "Biomimetic Concept Generation Applied to Design for Remanufacture," *2002 ASME Design Engineering Technology Conference and Company and Information in Engineering Conference*, Montreal, Quebec, Canada.
[15] McAdams, D. A., and Wood, K. L., 2000, "Quantitative Measures for Design by Analogy," , *DETC' 00, 2000 ASME Design Engineering Technology Conference* Baltimore, Maryland.
[16] Hirtz, J., Stone, R. B., and McAdams, D. A., 2002, "A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts," Res. Eng. Des., **13**, pp. 65–82.
[17] Gentner, D., and Markman, A. B., 1997, "Structure Mapping in Analogy and Similarity," Am. Psychol., **52**, pp. 45–56.
[18] Dahl, D. W., and Moreau, P., 2002, "The Influence and Value of Analogical Tthinking During New Product Ideation," J. Mark. Res., **39**(1), pp. 47–60.
[19] Wilson, J. O., Rosen, D., Nelson, B. A., and Yen, J., 2010, "The Effects of Biological Examples in Idea Generation," Des. Stud., **31**, pp. 169–186.
[20] Gick, M. L., and Holyoak, K. J., 1980, "Analogical Problem Solving," Cogn. Psychol., **12**(3), pp. 306–355.
[21] Dunbar, K., 1997, "How Scientists Think: On-Line Creativity and Conceptual Change in Science," *Creative thought: An investigation of conceptual structures and processes*, T. B. Ward, S. M. Smith, and J. Vaid, ed., Amer. Psych. Assoc., Washington, DC, pp. 461–493.
[22] Weisberg, R. W., 2009, "On "Out-of-the-Box," Thinking in Creativity," *Tools for Innovation* A. B. Markman and K. L. Wood, eds., Oxford University Press, New York.
[23] Purcell, A. T., and Gero, J. S. 1992, "Effects of Examples on the Results of a Design Activity," Knowledge-Based Syst., **5**(1), pp. 82–91.
[24] Duncker, K., 1945, *On Problem Solving*, Amer. Psych. Assoc. Washington, DC.
[25] Adamson, R. E., 1952, "Functional Fixedness as Related to Problem Solving: A Repetition of Three Experiments," J. Exp. Psychol., **44**(4), pp. 288–291.
[26] Maier, N. R. F., 1931, "Reasoning in Humans. II. The Solution of a Problem and Its Appearance in Consciousness," J. Comp. Psychol., **12**, pp. 181–194.
[27] Chase, W. G., and Simon, H. A., 1973, "The Mind's Eye in Chess," *Visual Information Processing*, W. G. Chase, ed., Academic Press, New York.
[28] Chi, M. T. H., Feltovich, P. J., and Glaser, R., 1981, "Categorization and Representation of Physics Problems by Experts and Novices," Cogn. Sci., **5**, pp. 121–152.
[29] Chi, M. T. H., and Koeske, R. D., 1983, "Network Representation of a Child's Dinosaur Knowledge," Dev. Psychol., **19**(1), pp. 29–39.
[30] Newell, A., 1990, *Unified Theories of Cognition*, Harvard University Press, Cambridge, MA.
[31] Anderson, J. R., and Schunn, C. D., 2000, "Implications of the ACT-R Learning Theory: No Magic Bullets," *Advances in Instructional Psychology*, R. Glaser, ed., Lawrence Erlbaum, Mahwah, NJ.
[32] Kaplan, C. A., and Simon, H. A., 1990, "In Search of Insight," Cogn. Psychol., **22**, pp. 374–419.
[33] Ohlsson, S., 1992, "Information-Processing Explanations of Insight and Related Phenomena," *Advances in the Psychology of Thinking*, Vol. **1**, M. T. Keane and K. J. Gilhooly, eds., Harvester Wheatsheaf, Hertfordshire, UK.
[34] Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D., 1999, "Constraint Relaxation and Chunk Decomposition in Insight Problem Solving," J. Exp. Psych. Learn. Mem. Cogn., **25**(6), pp. 1534–1555.
[35] McKoy, F. L., Vargas-Hernandez, N., Summers, J. D., and Shah, J. J., 2001, "Influence of Design Representation on Effectiveness of Idea Generation," *DETC '01: ASME 2001 Des. Eng. Tech. Conf. and Comp. and Inf. In Eng. Conf.*, Pittsburgh, PA.
[36] Green, M., Dutson, A., Wood, K. L., Stone, R., and McAdams, D., 2002, "Integrating Service-Oriented Design Projects in the Engineering Curriculum," *Proceedings of the 2002 American Society for Engineering Education Annual Conference and Exposition*.
[37] Green, M., and Wood, K. L., 2004, "Service-Learning Approaches to International Humanitarian Design Projects: Assessment of Spiritual Impact," *Proceedings of the 2004 Christian Engineering Education Conference*.
[38] White, C., and Wood, K. L., 2010, "Influences and Interests in Humanitarian Engineering," *Proceedings of the ASEE Annual Conference*, Lexington, KY, June 2010, AC 2010-652; *Proceedings of the Global Colloquium on Engineering Education*, Singapore, October 2010.
[39] Moss, J., Kotovsky, K., and Cagan, J., 2007, "The Influence of Open Goals on the Acquisition of Problem Relevant Information," J. Exp. Psych. Learn. Mem. Cogn. **33**(5), pp. 876–891.
[40] 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**, pp. 203–221.
[41] Markman, A. B., and Wood, K. L., eds., 2009, *Tools for Innovation: The Science Behind Practical Methods That Drive New Ideas*, Oxford University Press, New York.
[42] Boden, M. A., 2004, *The Creative Mind: Myths and Mechanisms*, 2nd ed., Routledge, London.
[43] Shah, J. J., Vargas-Hernandez, N., and Smith, S. M., 2003, "Metrics for Measuring Ideation Effectiveness," Des. Stud., **24**, pp. 111–134.
[44] Girotra, K., Terwiesch, C., and Ulrich, K. T., 2010, "Idea Generation and the Quality of the Best Idea," Manage. Sci., **56**(4), pp. 591–605.
[45] Blanchette, I., and Dunbar, K., 2000, "How Analogies are Generated: The Roles of Structural and Superficial Similarity," Mem. Cognit., **28**(1), pp. 108–124.
[46] Ishibashi, K., and Okada, T., 2006, "Exploring the Effect of Copying Incomprehensible Exemplars on Creative Drawings," R. Sun, ed., *Proceedings 28th Ann. Conf. Cog. Sci. Society*. Vancouver, Canada.
[47] Novick, L. R., 1988, "Analogical Transfer, Problem Similarity, and Expertise," J. Exp. Psych. Learn. Mem. Cogn., **14**(3), pp. 510–520.
[48] Chakrabarti, A., Sarkar, P., Leelavathamma, B., and Nataraju, B. S., 2005, "A Functional Representation for Biomimetic and Artificial Inspiration of New Ideas," *AIEDAM*, **19**, pp. 113–132.
[49] Forbus, K. D., Gentner, D., and Law, K., 1994, "MAC/FAC: A Model of Similarity-Based Retrieval," Cogn. Sci., **19**, pp. 141–205.