Chapter 12
Do the Best Design Ideas (Really) Come from Conceptually Distant Sources of Inspiration?

Joel Chan, Steven P. Dow and Christian D. Schunn

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

12.1 Background

12.1.1 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,

J. Chan (✉) · C. D. Schunn
Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: joc59@pitt.edu

J. Chan
School of Information Studies, University of Maryland, College Park, Maryland, USA

S. P. Dow
Human–Computer-Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA

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 reusable 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 and Markman 1997; Holyoak and 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 and Golden 1997), to more recent case studies (Enkel and Gassmann 2010; Kalogerakis et al. 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 and Shu 2012; Dahl and Moreau 2002; Gonçalves et al. 2013; Hender et al. 2002); but, some in vivo studies of creative cognition have not found strong connections between far sources and creative mental leaps (Chan and Schunn 2014; Dunbar 1997), and other experiments have demonstrated equivalent benefits of far and near sources (Enkel and Gassman 2010; Malaga 2000). Relatedly, Tseng et al. (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 et al. (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.

12.1.2 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 and 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 and Smolkov 2006; Liikkanen and Perttula 2008).

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) reexamining theories of design inspiration and conceptual distance in light of accumulating preponderance of evidence *against* the Conceptual Leap Hypothesis.

12.2   Methods

12.2.1   Overview of Research Context

The current work is conducted in the context of OpenIDEO (www.openideo.com), a Web-based crowdsourced 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 12.1 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

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.

Fig. 12.1 Example concept illustrating the typical amount of detail per concept

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 also able to see concepts/inspirations the given concept/inspiration "built upon" (i.e., cited as explicit sources of inspiration; see Fig. 12.2). This culture of citing sources is particularly advantageous, given that people generally forget to monitor or cite their sources of inspiration (Brown and Murphy 1989; Marsh et al. 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 and Dunbar 1996).
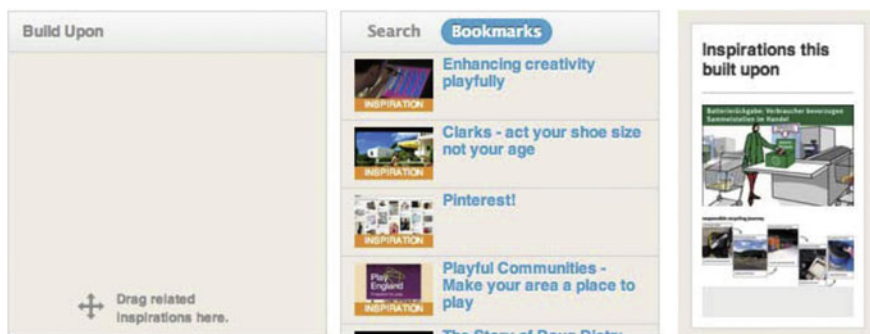

Fig. 12.2 Depiction of OpenIDEO citation workflow. When posting concepts/inspirations, users are prompted to cite concepts/inspirations they "build upon" by dragging bookmarked concepts/inspirations (middle panel) to the citation area (left panel). Users can also search for related concepts/inspirations at this step (middle panel). These cited sources then show up as metadata for the concept/inspiration (right panel)

12.2.2 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 12.2 for a description of the 12 challenges (with some basic metadata on each challenge). Figure 12.3 shows the full-text design brief for two challenges.

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.

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.
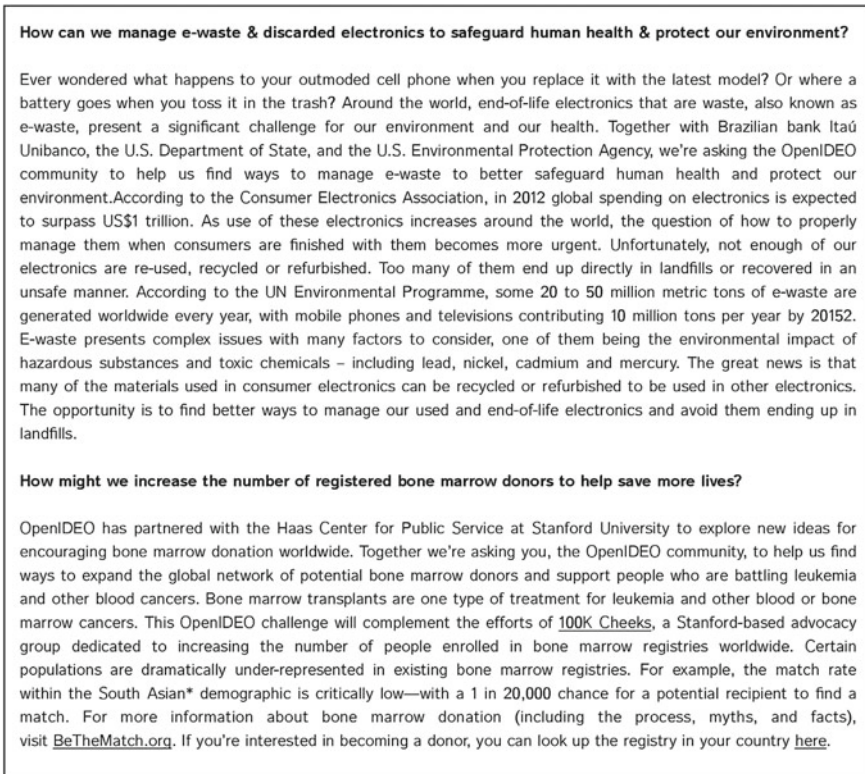
Fig. 12.3 Full-text of challenge briefs from two OpenIDEO challenges

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 ($\sim$16%) were shortlisted (see Table 12.1 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 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.

Table 12.1 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 and discarded electronics to safeguard human health and 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)

We analyze the impact of the distance of inspirations (and not cited concepts) given our focus on ideation processes during "original" or nonroutine 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 nonroutine design.

12.2.3 Measures

12.2.3.1 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 and McKool 2009; Brown 1989; Sawyer 2012). Further, we know from conversations with the OpenIDEO team that the panel's judgments combine 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) the 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).

12.2.3.2 Conceptual Distance

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 and 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 et al. 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 and Steyvers 2004), describing the structure and evolution of scientific fields (Blei and Lafferty 2006, 2007), and discovering topical dynamics in social media use (Schwartz et al. 2013).

We use Latent Dirichlet Allocation (LDA; Blei et al. 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.

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 1. 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 and Martin 2001).

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 on 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 Fig. 12.4A). 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) 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 Fig. 12.4b), further validating the LDA approach to measuring distance. Figure 12.5 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
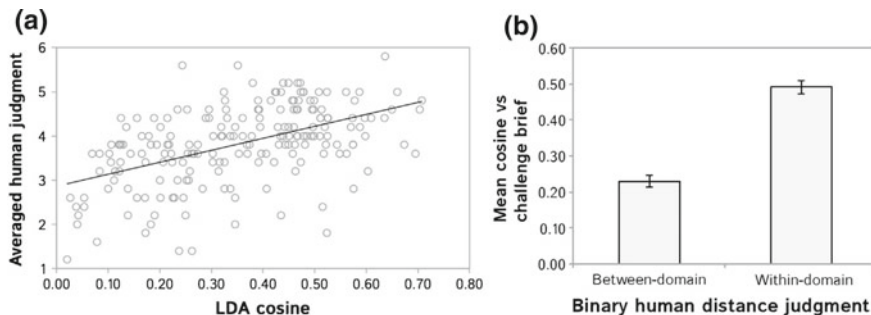
Fig. 12.4 **a** Scatterplot of LDA cosines vs. averaged human continuous similarity judgments for inspirations in the e-waste challenge. **b** Mean cosine against the challenge brief for within—versus between-domain inspirations

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 prior effort to address e-waste.

The near and far examples depicted in Fig. 12.5 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 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).

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
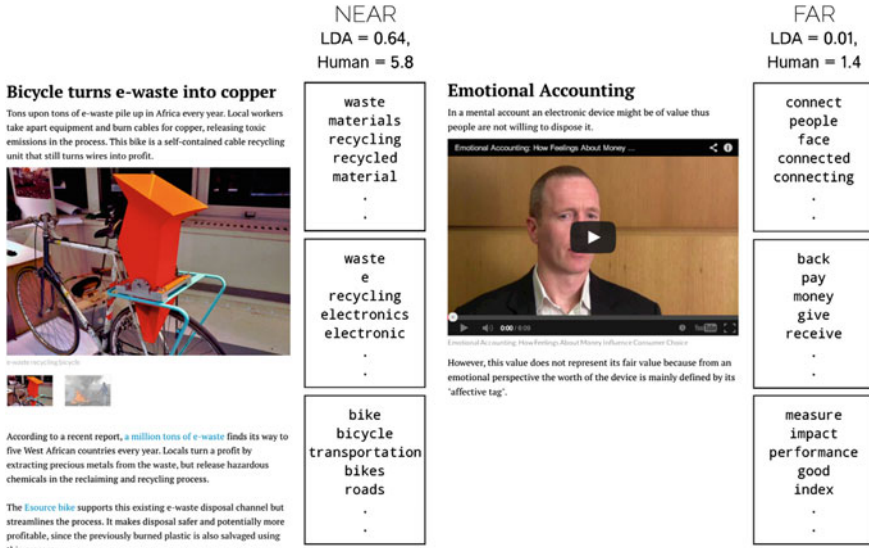
Fig. 12.5 Topics found by LDA within examples of near and far inspirations for the e-waste challenge

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 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).

12.2.3.3 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.

12.2.4   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 Fig. 12.6). 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 nonindependence could lead to overestimates of the statistical significance of model estimates (i.e., make unwarranted claims of statistically significant effects). This issue is exacerbated
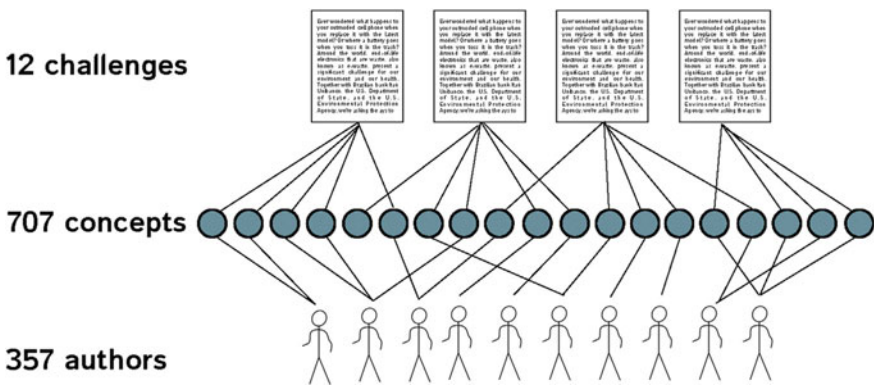
Fig. 12.6 Illustrated cross-classified structure of the data

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 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]

12.3 Results

12.3.1 Descriptive Statistics

On average, 16% of concepts in the sample get shortlisted (see Table 12.2). $DIST_{MEAN}$ 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 12.3). 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.

12.3.2 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*.

---

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

Table 12.2 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

Table 12.3 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**

$^m p < .10$, $*p < .05$, $**p < .01$, $***p < .001$

12.3.2.1 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 2 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 12.7 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 to 0.11]; using a logistic model, the coefficient for "any citation" (controlling for feedback) is 0.31, 95% CI = [0.01 to 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 advantage of citing versus 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
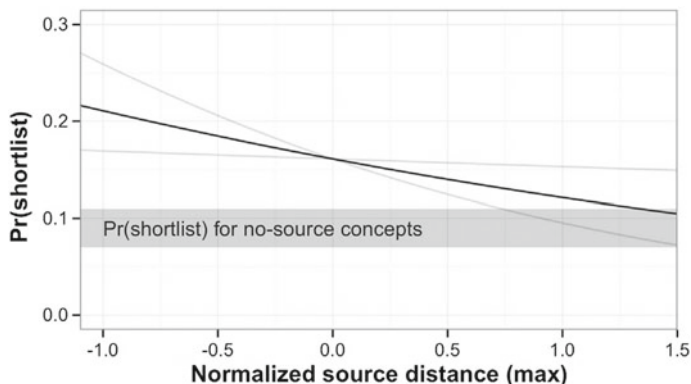
Fig. 12.7 Model-fitted relationship between $DIST_{MAX}$ and Pr (shortlist), evaluated at mean values of feedback and source shortlist. Grayed lines are fits with upper and lower limits for 95% CI for effect of $DIST_{MAX}$
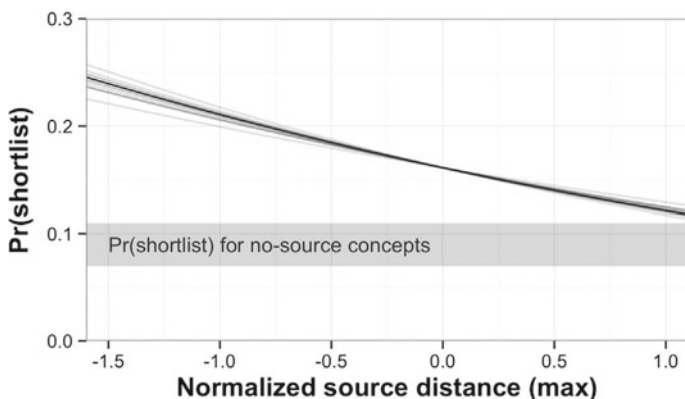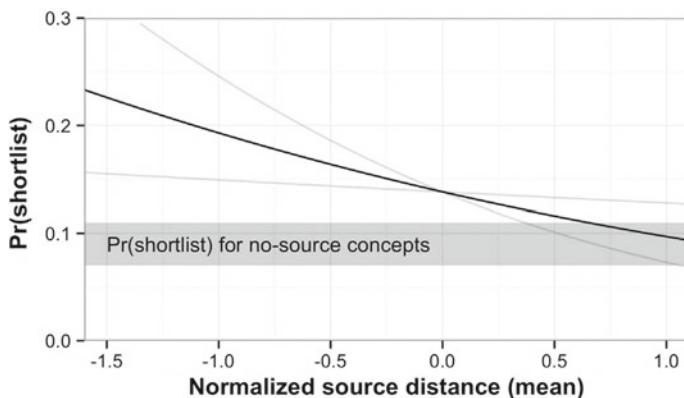

Fig. 12.8 Overall and by-challenge model-fitted relationship between $DIST_{MAX}$ and Pr (shortlist). Fitted values evaluated at mean values of feedback and source shortlist. Grayed lines are fits for each individual challenge

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 $DIST_{MAX}$. $\chi^2(2) = 0.05$, $p = 0.49$ (see Fig. 12.8).

12.3.2.2   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 short-listed, $p < .05$. 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 Fig. 12.9, 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 Fig. 12.10).

12.4   Discussion

12.4.1   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 contrary, the benefits of building

Fig. 12.9 Model-fitted relationship between $DIST_{MEAN}$ 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 $DIST_{MEAN}$
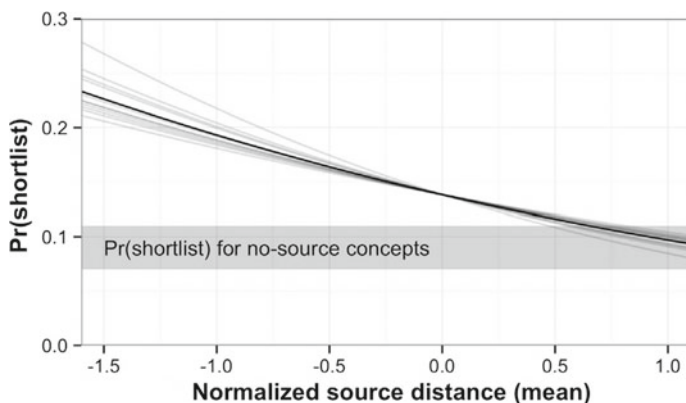
Fig. 12.10 Overall and by-challenge model-fitted relationship between $DIST_{MEAN}$ and Pr (shortlist). Fitted values evaluated at mean values of feedback and source shortlist. Grayed lines are fits for each individual challenge

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 nonroutine social innovation design problems.

12.4.2 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 underreported). 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, 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 et al. 2013). 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).

12.4.3 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 and Schunn 2014; Dow et al. 2009; Mecca and Mumford 2013; Rietzschel et al. 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 (Hargadon and Sutton 1997; Franke et al. 2013; Jeppesen 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 (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 nonexistent.

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 (Knoblich et al. 1999; Kaplan and Simon 1990; Öllinger et al. 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 and 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 and Christensen 2009; Yu et al. 2014). Are there processing strategies that expert creative designers apply uniquely to far sources (e.g., to deal with potentially un-alignable 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 1: Topic Model Technical Details

Document Preprocessing

All documents were first tokenized using the TreeBank Tokenizer from the open-source Natural Language Tool Kit Python library (Bird et al. 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.

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 et al. 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 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 et al. 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 12.11 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. Since 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
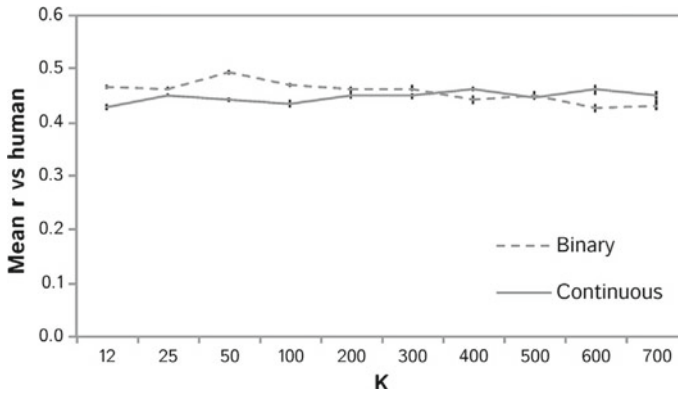
Fig. 12.11 Mean fit (with ± 1 SE) versus human judgments for LDA cosines by level of K

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.

Appendix 2: Statistical Modeling Technical Details

Statistical Modeling Approach

All models were fitted using the lme4 package (Bates et al. 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):

$$\eta_{i(\text{authorj challengek})} = \gamma_{00} + \sum_q \gamma_{q0} X_{qi} + u_{0\text{authorj}} + u_{0\text{challengek}}$$

where

- $\eta_{i(\text{authorj challengek})}$ is the predicted log-odds of being shortlisted for the ith concept posted by the jth author in the kth 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)
- $u_{0\text{authorj}}$ and $u_{0\text{challengek}}$ 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 $DIST_{\text{MAX}}$ and $DIST_{\text{MEAN}}$, 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 $u_{1\text{challengek}}$ that models the between-challenge variance component for the slope of distance.

Model Selection

Estimates and test statistics for each step in our model-building procedure are shown in Tables 12.3 and 12.4. 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, $p = 0.21$; 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, $\chi^2(2) = 0.13$, $p = .47$.

Table 12.4 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			
$u_{0\text{authorj}}$ for intercept	0.29	0.31	0.32
$u_{0\text{challengek}}$ for intercept	0.75	0.76	0.74
$u_{3\text{challengek}}$ 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]

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 and 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 versus null at $\chi^2(3) = 79.71$, $p = .00$.

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 $\chi^2(3) = 80.61$, $p = .00$ (Table 12.5).

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

Table 12.5 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)	$DIST_{MEAN}$, fixed slope	$DIST_{MEAN}$, 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]

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

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

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.

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). PA: Hershey.

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*. Sebastopol, CA: 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, 1,* 17–35.

Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3,* 993–1022.

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

Brown, R. (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: Plenum Press.

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. https://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., & Schunn, C. (2014). The impact of analogies on creative concept generation: Lessons from an in vivo study in engineering design. *Cognitive Science, 39*(1), 126–155.

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. In *Advances in neural information processing systems* (pp. 288–296).

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.

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

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, DC: American Psychological Association Press.

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*. Thousand Oaks, CA: Sage.

Franke, N., Poetz, M. K., & Schreier, M. (2013). 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: 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. https://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*(ahead-of-print), 1–18.

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. https://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. https://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: MIT press.

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), 1016e1033.

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

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

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.

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), 0410031–04100312.

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 and 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, 48*(3), 209–236. https://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, 39*(3), 931. https://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*. Berlin: 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.). CA: Thousand Oaks.

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.

Sawyer, R. K. (2012). *Explaining creativity: The science of human innovation* (2nd ed.). New York: Oxford University Press.

Schwartz, H. A., Eichstaedt, J. C., Kern, J. K., Dziurzynski, J. K., Ramones, M. L. D., Agrawal, L. R., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE, 8*(9), e73791.

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

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* (3rd ed.). New York: McGraw Hill.

Wallach, H. M., Mimno, D. M., & McCallum, A. (2009). Rethinking LDA: Why priors matter. *Neural Information Processing Systems, 22,* 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). Bulgaria: Sofia.

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). NY: New York.

Weisberg, R. W. (2011). Frank lloyd wright's fallingwater: A case study in inside-the-box creativity. *Creativity Research Journal, 23*(4), 296–312. https://doi.org/10.1080/10400419.2011.621814.

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory and 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., and 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, 44,* 1049–1060.