THE IMPACT OF SOURCES OF INSPIRATION ON THE GENESIS OF INNOVATIVE IDEAS

by

Chu Sern Joel Chan

Bachelor of Science, University of the Ozarks, 2009

Master of Science, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

The Kenneth P. Dietrich School of Arts and Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Chu Sern Joel Chan

It was defended on

July 1, 2014

and approved by

Timothy J. Nokes-Malach, Associate Professor, Psychology, Learning Sciences and Policy

Kevin D. Ashley, Professor, Law, Intelligent Systems

Steven P. Dow, Assistant Professor, Human-Computer Interaction, Carnegie Mellon University

Dissertation Advisor: Christian D. Schunn, Professor, Psychology, Learning Sciences and Policy, Intelligent Systems

THE IMPACT OF SOURCES OF INSPIRATION

ON THE GENESIS OF INNOVATIVE IDEAS

Chu Sern Joel Chan, PhD

University of Pittsburgh, 2014

Innovation fundamentally begins with a good idea. But where do good ideas come from? Much research suggests that innovative breakthroughs are often inspired by past experience: things and ideas that one has interacted with in the world. However, the same experiences that can inspire innovation can sometimes constrain or harm innovation through focus on previously unsuccessful solutions. In this dissertation, I explore principles for guiding interactions with sources of inspiration (previous/other ideas) to maximize their benefits and minimize their pitfalls, focusing on the role of conceptual *distance* and *diversity* of sources. I analyze thousands of ideas for complex innovation challenges (e.g., increasing accessibility in elections, revitalizing struggling urban areas) posted to an online crowd-sourced innovation platform that required contributors to cite sources of ideas, tracing the impact of the distance and diversity of sources in ideas' conceptual genealogies on their creative success (as judged by an expert panel).

In this dissertation, I make three primary contributions to the literature. First, leveraging techniques from natural language processing and machine learning, I develop a validated computational methodology for studying conceptual distance and diversity with complex design concepts, which addresses significant issues of efficiency and scalability faced in prior work. Second, I challenge the widespread but unevenly supported notion that far sources provide the best insights for creative ideation; addressing key methodological issues in prior work (time scale, statistical power, and problem variation), I show that overreliance on far sources can harm ideation success, and that good ideas can often come from very near sources. Finally, I demonstrate the potential value of incorporating a temporal dimension into analyses of the impact of sources of inspiration: I find evidence of differential impacts of source distance and diversity (viz., increased problem variation for the effect of source distance, and a more robust positive effect of source diversity) when considering sources farther back in ideas' conceptual genealogies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

It is hard to believe that 5 years have already passed, and I have passed through some invisible gate, emerging as an equal partner in the world of scientific research. This transition has been gradual and, thankfully, now feels altogether fitting: standing on the shoulders of giants, I feel ready for (God-willing) a lifetime of inquiry and contributions to the science of cognition. I would be remiss, however, if I did not acknowledge at least some of the many giants on whose shoulders I now stand. Foremost among these giants is Chris Schunn, my advisor. Fearless, endlessly energetic and creative, with one of the sharpest and quickest minds I have been privileged to interact with, you have taught me so much about *real*, creative, interdisciplinary inquiry. Without your expert guidance and ever-present faith in my potential, I would not where I am today. I hope to one day begin to match your research brilliance and fruitful mentorship. The next group of giants is comprised of my dissertation committee (Steven Dow, Kevin Ashley and Timothy Nokes-Malach), my mentoring committee (Kevin and Tim again!), the Higher-Order Cognitive Collective (particularly Susannah Paletz, Kevin Soo, Meghan Bathgate, Liz Richey, JooYoung Jang, Carmela Rizzo, Cristina Zepeda, Dan Belenky, Matt Bernacki, Lou Alfieri, Amanda Crowell, Matty Lau, Sam Abramovich, and Melissa Patchan), and the venerable Verrocchio research group that launched me on this crazy path of studying engineering innovation (Kate Fu, Jon Cagan, Ken Kotovsky, and Kris Wood). From these wonderful individuals I have derived an endless supply of inspiration, constructive criticism, research ideas, and encouragement. Many of the ideas represented in this dissertation have been forged in the fire of collaboration with these wonderful people. A third group of giants are those who helped form me as a lifelong learner during my undergraduate education, both at university and in my research internship: Tom Stepehenson, Karen Jones, Joel Hagaman, Mikael Lindstrom, Sharon Gorman, Gilbert Parks, Greg Trafton, and Raj Ratwani. A fourth group of giants deserves mention: my extended adopted family in Pittsburgh from City Reformed Presbyterian Church, who have tirelessly helped to keep me grounded in the joy of where I am, and to maintain a healthy work-life balance as I discover the fingerprints of God in the human mind. Last, but certainly not least, my family – my bedrock of support, unconditional love, and continual guidance, encouragement, and inspiration — and Anna, my lovely wife: I am so blessed to walk this road with you, and I am excited for what's next!

1.0 GENERAL INTRODUCTION

1.1 MOTIVATION AND OVERVIEW OF INQUIRY CONTEXT

Creativity and innovation are crucial mainstays in modern society. Continued innovation is a central driver of today's knowledge-based economy; in order to survive and thrive, firms can no longer depend on commoditization and scale — they must innovate or die (Vogel, Cagan, & Boatwright, 2005). The U.S., too, needs innovation to continue thriving in an increasingly globalized and knowledge-driven economy (National Academy of Engineering, 2005). Further, complex problems facing modern society, such as global poverty, e-waste, cancer, and climate change, are more pressing than they have ever been, and call for new innovative solutions. How can governments, organizations, and training programs effectively *train* and *support* innovators to rise to these challenges? A crucial part of the puzzle is a robust scientific knowledge base that articulates key principles of how creativity happens. Cognitive science offers a key piece of this knowledge base, focusing on the creative process (including mental processes and strategies) that lead to creative breakthroughs.

One of the most robust and established insights from the cognitive science of creativity is that the creation of new ideas is strongly constrained or structured by prior knowledge and experience. People have a strong tendency to transfer features and elements from recently encountered stimuli or examples in their creative production, often despite instructions to avoid such copying (Jansson & Smith, 1991; Marsh, Bink, & Hicks, 1999; Marsh, Ward, & Landau, 1999; Perttula & Sipilä, 2007; Purcell & Gero, 1992, 1996; Smith, Ward, & Schumacher, 1993; Ward, 1994). This tendency towards transfer can harm creativity. Some experiments have shown that people will transfer elements from examples even when those examples are known to be of low quality (Chrysikou & Weisberg, 2005; Jansson & Smith, 1991). Prior knowledge can also lead to functional fixedness — the inability to see novel uses for an artifact due to prior conceptions of its dominant functionality (Adamson, 1952; German & Barrett, 2005; Maier, 1931) — and mental set effects (also called Einstellung) — where people persist in using previously successful solution approaches in problem solving without considering alternative, potentially more effective, approaches for the current task at hand (Bilalić, McLeod, & Gobet, 2008; Luchins, 1942; Öllinger, Jones, & Knoblich, 2008; Wiley, 1998).

However, this tendency to base creative production on prior knowledge is not an inherent detractor from creativity. Purcell and Gero (1996) have argued that fixation is an imprecise (and perhaps incorrect) term for this phenomenon: when the examples are innovative or of high quality, the transfer may increase the creativity of the final product. Conformity to examples does not necessarily influence other key process measures of creativity, such as quantity or elaboration of concepts generated (Marsh, Landau, & Hicks, 1996), and, depending on features of the source (e.g., novelty, conceptual distance from domain), may also increase the quality of ideas (Ward, 2008), or also novelty of ideas (Chan et al., 2011; Smith, Kohn, & Shah, 2008).

For these reasons, intentional curation of the building blocks of prior knowledge/experience—hereafter called **sources of inspiration**—is a crucial component of effective creative practice. Tradecraft literature (e.g., books, blogs) is replete with advice and support for curating sources of inspiration: Henry (2011) urges creators to carefully curate stimuli to keep their creative fuel burning, and Dyer and colleagues (2011) urge innovators to keep their "pool of available bricks" in memory fresh, to increase the probability that truly breakthrough concepts can be generated. Detailed ethnographic studies of successful innovators and creators have also corroborated the central role of curating and intentionally interacting with sources of inspiration (Eckert & Stacey, 1998; Hargadon & Sutton, 1997; Herring et al., 2009). Further, the issue of how to prevent and/or alleviate design fixation is an active area of research in design methodology research (Linsey et al., 2010; Youmans, 2011; Zahner et al., 2010).

But *how* should one curate one's sources of inspiration? Or to pose the question more precisely, **what principles should guide the curation and use of sources of inspiration in the creative process such that creators can benefit from them while avoiding their potential pitfalls?** One key facet of this question concerns the nature of the sources themselves: are there particular features or properties of inspirational sources (e.g., conceptual distance to the problem, conceptual diversity among considered sources) that provide reliable signals of greater or lesser inspirational potential?

This dissertation addresses these fundamental questions with quantitative analyses of creative processes and outputs of individuals solving real-world creative design problems, focusing on the issue of conceptual distance. I focus on this issue given the discrepancy between the widespread claims offered in the scientific literature and among practitioners as to how conceptual distance of and between sources can matter for creative outcomes, and the strength (or lack thereof) of the evidence base for these claims. This presents an opportunity for significant knowledge gains to be made, in contrast to other relatively uncontroversial claims regarding the nature of sources (e.g., build on high-quality solutions).

The aim of this dissertation is to yield insights into the specific question of what principles should guide the curation of inspiration sources, and also more generally for efforts to understand and maximize creativity and innovation, from the design and implementation of innovation support tools and methods (e.g., computer-aided design, formal design-by-analogy methods), to the new wave of creative crowdsourcing platforms (similar to OpenIDEO), to creativity education in the disciplines, to the intentional design of creative social spaces (e.g., R&D centers, innovation hubs).

1.2 CONCEPTUAL DISTANCE

1.2.1 Overview

The first major line of inquiry in this dissertation examines the role of the conceptual distance of sources. 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 a source that is conceptually near to the problem domain, 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?

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 innovative 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 innovative 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).

1.2.2 Research Base and Opportunities for Advancement

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 innovation 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; Tseng, Moss, Cagan, & Kotovsky, 2008), and even harmful effects of distance (Fu et al., 2013). Thus, more empirical work is needed to determine whether the Conceptual Leap Hypothesis is well supported.

Key methodological shortcomings in prior work further motivate more and better empirical work. First is the issue of *time scale.* Prior studies may be too short (typically 30 minutes to 1 hour) to convert far sources into viable concepts. Scarce cognitive resources are required 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).

1.3 CONCEPTUAL DIVERSITY

1.3.1 Overview

The second major line of inquiry in this dissertation considers the hypothesis that, in using sources of inspiration, one should attempt to connect sources and concepts that are conceptually far from each other. We shall call this the Conceptual Combination Hypothesis. In the course of a concept's development, designers often build on ideas from more than one source (e.g., different approaches for a single sub-system, different sources for different sub-systems). Consider again an innovator developing creative solutions for the problem of e-waste accumulation. She might build on related but slightly different approaches to educating about e-waste, such as classroom curricula, video education series on Youtube, and on-label information about reuse/recycle options (near combinations); alternatively, she might combine concepts from gamification, social media campaigning and marketing, and exercise and dieting lifestyle-change mobile applications (far combinations). The hypothesis being investigated is that a breakthrough creative solution is more likely in the latter case.

This hypothesis is related to the Conceptual Leap Hypothesis, but distinct in that it does not necessarily distinguish between combining sources that are far from each other within the problem domain (e.g., combining a bus and a plane to come up with a new transportation system) and far combinations from within to outside (e.g., combining a bicycle and a printer), or

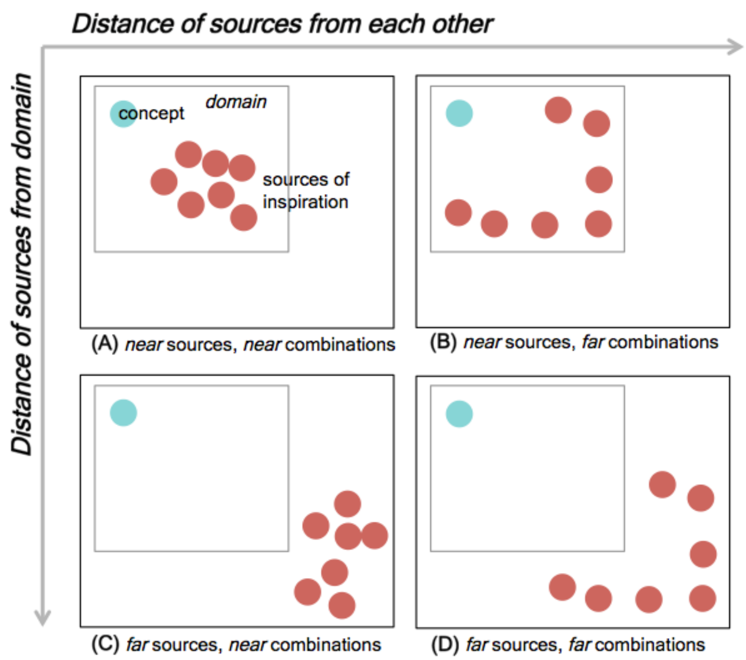

Figure 1: Illustrated variations of inspiration source sets

to sources outside the problem domain (e.g., combining a heart defibrillator with geese migration patterns). **Figure 1** illustrates the range of possible variations in source sets by distance from domain and distance of combination. It is important to understand not just how each dimension of conceptual distance influences ideation separately, but also how they might interact.

1.3.2 Research Base and Opportunities for Advancement

The recommendation to prefer far combinations has its scientific roots in Mednick's (1962) influential claim that "[t]he more mutually remote the elements of the new combination, the more creative the process or solution" (p. 221). Further, theorists who analyze tecnological innovation through studying patent citation networks contend that patents that reference other patents from a wide number of technology areas hold more potential for radical innovation compared to patents that reference other patents in similar technology areas (Olsson, 2005). Recent social network theories of innovation have similarly emphasized the importance of combining information from diverse sources as a basis for innovation (Vedres & Stark, 2010). These theoretical ideas are consistent with many anecdotes of creative breakthroughs coming from far combinations, such as award-winning chef Samuelsson's fusion Swedish cuisine, Grammy-winning singer-songwriter Shakira's fusion of Latin and hip-hop musical sounds, and the highly successful Magic collectible cards game, which combined concepts from collectible items (such as baseball trading cards) and ordinary games (Johansson, 2006).

Far conceptual combinations might support creative breakthroughs via the generation of emergent features when trying to combine them. Research on conceptual combination suggests that, when concepts are very different, people switch from relatively simpler combination processes — such as attribute inheritance/transfer or property mapping (Hampton, 1987; Wisniewski & Gentner, 1991) — to more complex processes, such as structure mapping (Gentner et al., 1997), which can generate emergent features (attributes that are true of neither constituent, but true of the conjunction; Hampton, 1997). Relatedly, the degree to which emergent features arise from combinations has been found to be an inverse function of the conceptual similarity between the constituent concepts (Wilkenfeld, 1995; Wilkenfeld & Ward, 2001; Wisniewski, 1997). Thus, combining concepts that are conceptually far from each other is likely to result in original features and functions that might prove to be crucial components of a creative breakthrough.

Far combinations may also protect against fixation: considering sources far from each other in conceptual space may prevent one from getting too strongly stuck in one region of the conceptual space, perhaps due in part to the distribution of memory activation across a wider range of features and functions. Another possible inspirational mechanism of far combinations (or at least having a diverse set of sources to draw from) is the increased statistical probability of finding an interesting and potentially useful combination (Simonton, 1988), although this mechanism might only operate if the set consists mostly of useful rather than completely irrelevant sources.

Experimental and observational studies generally support the Conceptual Combination Hypothesis, although much of it focuses on the effects of far combinations on *novelty* of ideas generated. The ability to create high-quality and original emergent features from conceptual combinations has been associated with performance on creative problem-solving tasks (Mumford, Baughman, & Sager, 2003). Generating ideas using stimuli from different categories have been shown to yield more novel ideas than using stimuli from similar (or the same) categories, both in simple brainstorming experiments with toy problems (Baughman & Mumford, 1995; Howard-Jones, Blakemore, Samuel, Summers, & Claxton, 2005; Zeng, Proctor, & Salvendy, 2011), with more realistic creative tasks like graphic design or business opportunity identification (Chase, Herman, & Dow, 2012; Gielnik, Frese, Graf, & Kampschulte, 2011), although sometimes at the expense of idea quality (Mobley, Doares, & Mumford, 1992). Baruah

(2011) found no positive effect on originality, but did find a positive effect on breadth of search, with participants who were stimulated with distantly related categories surveying more idea categories than participants stimulated with closely related categories. Nijstad and colleagues (2002) demonstrated similar benefits of stimuli diversity on breadth of search. In a more ecologically valid setting, Taylor and Greve (2006) showed that comic book creators' diversity of prior genre experience positively predicted creative performance (measured in terms of collector market value of comics produced). In general, the literature provides support for a link between distant combinations and the novelty of ideas generated, but most studies (except Taylor & Greve, 2006) have not yet looked carefully at whether there is an effect on final quality of ideas.

1.4 RESEARCH QUESTIONS AND OVERVIEW OF DOCUMENT

In this dissertation, I contribute to knowledge on curation of inspiration sources by testing the Conceptual Leap and Conceptual Combination Hypotheses, addressing key methodological issues in prior work (e.g., time scale, problem variation, novelty vs quality in creative outcomes). Specifically, the two primary research questions addressed in this dissertation are:

1) **What are the relative benefits of different levels of source conceptual distance for creative outcomes?**
2) **What are the relative benefits of different levels of source conceptual combination distance for creative outcomes?**

I examine these questions in the context of OpenIDEO (www.openideo.com), a large-scale Web-based crowd-sourced innovation platform where thousands of individuals have been coming together to collaboratively solve a wide range of socially and environmentally important problems (e.g., managing e-waste, increasing accessibility in elections, restoring community in socially fragmented cities). Contributors to the platform follow a structured design process — starting from initial problem structuring, through concept generation and screening, to refinement and evaluation of concepts — to produce concepts that are ultimately implemented by the challenge sponsors, producing real-world impact. I trace how variations in conceptual distance of sources from the problem domain, and conceptual distance *among* sources, relate to creative success (i.e., the creation of designs that are both novel and add significant value over existing designs).

The remainder of this document consists of six remaining chapters. In Chapter 2, I describe in more detail the overall research context and methodological approach. Chapters 3-5 examine three different angles on the issue of conceptual distance: chapter 3 examines the role of conceptual distance from one's problem domain; chapter 4 examines the role of conceptual distance from one's solution path; and chapter 5 examines the related notion of conceptual distance *between* one's sources. Then, in Chapters 6 through 8, I leverage the rich structure of my data to explore more fine-grained variations of the main research questions, specifically exploring how the effects of distance and/or diversity might be different for "indirect" sources (i.e., sources of one's immediate sources, or more informally, one's "conceptual genealogy"). Finally, Chapter 9 integrates the insights yielded from this dissertation, and examines implications for the theory and practice of creative inspiration, and opportunities for further research.

2.0 GENERAL METHODS

2.1 OVERVIEW OF RESEARCH CONTEXT

OpenIDEO (www.openideo.com) is 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 (see **Table 1**) to produce concepts that are ultimately implemented for real-world impact ("Impact Stories," n.d.). The overall analysis focuses on two crucial early stages in the process: first, in the *inspiration* phase (lasting between 1.5 to 4 weeks, M = 3.1), contributors post *inspirations* (e.g., descriptions of solutions to analogous problems, case studies of stakeholders), which help to define the problem space and identify promising solution approaches. The OpenIDEO designers guide this inspiration phase by soliciting specific kinds of inspirations, through “assignments”: some assignments call for descriptions of related efforts (e.g., battery collection initiatives, for solving the problem of e-waste); some call for interviews with stakeholders (e.g., how do users feel about their electronics?); while others explicitly solicit “far inspirations” (e.g., thinking of other situations in which lack of knowledge is a barrier to action). Then, in the *concepting* phase (lasting the next 2 to 6 weeks, m = 3.4), contributors post *concepts*, i.e., specific solutions to the problem. They are different from inspirations in that they are explicit, concrete proposals for how

Table 1: OpenIDEO structured design process

Phase	Description
0: Start	Community receives challenge brief; problem broadly framed; initial constraints/requirements described
1: Inspiration	Community submits, “applauds” (i.e., votes on), and gives feedback on inspirations (e.g., descriptions of solutions to analogous problems, case studies of stakeholders); problem space defined in more detail, promising solution approaches (“themes”) identified by administrators/sponsors
2: Concepting	Community submits, applauds, and gives feedback on concepts (proposed solutions to problem)
3: Screening	Using applause as input, administrators & sponsors *shortlist* subset of concepts for further refinement
4: Refinement	Community collaborates with authors to improve shortlisted concepts
5: Evaluation	Community provides focused evaluations of shortlisted concepts based on administrator & sponsor-defined challenge-specific evaluation rubrics
6: Realization	Administrators & sponsors select winning concepts for implementation

to solve the specific problem posed by the challenge, as opposed to information about the problem, descriptions of solutions to other problems, or vague descriptions of potential “entry points” for successful solutions (e.g., proposing a specific education plan for reusing electronics [*concept*], vs identifying lack of knowledge as a major barrier [*inspiration*]). In later stages, concepts are selected, refined, and implemented. **Figure 2** 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.

The OpenIDEO platform has many desirable properties as a research context for this work, including the existence of multiple design problems (22 as of February 2014), thousands

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 2: Example concept illustrating the typical amount of detail per concept

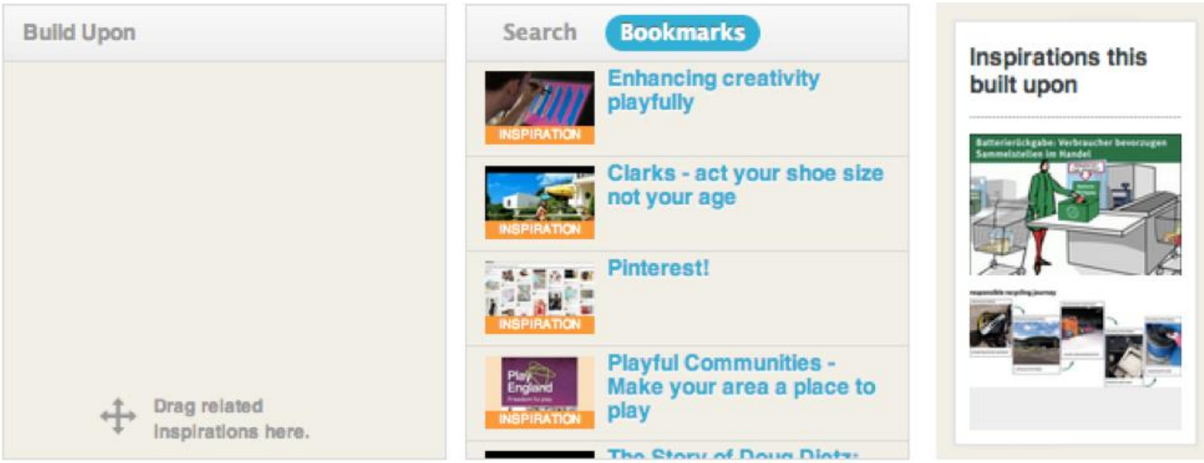

Figure 3: Depiction of OpenIDEO citation workflow. When posting concepts/inspirations, users are prompted to cite concepts/inspirations they "build upon" by dragging bookmarked concepts/inspirations (middle panel) to the citation area (left panel). Users can also search for related concepts/inspirations at this step (middle panel). These cited sources then show up as metadata for the concept/inspiration (right panel).

of concepts and inspirations, text-based record of ideas to enable efficient text-based analyses, and a record of feedback received, another critical factor in design success.

A central property for the research questions in this dissertation 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. Specifically, on the interface page where users post a concept there is a prominent interface for citing concepts (see **Figure 3**), with the following instructions: "Did someone else's Inspiration or Concept inspire your Concept? Drag across any contribution that did - you'll help everyone understand where yours came from and give you and the other user extra DQ points!" Here, "DQ points" refer to “design quotient” points that OpenIDEO users can receive and display on their profile for various actions on the platform (e.g., posting inspirations, concepts, commenting and/or citing/collaborating on other inspirations/concepts): importantly, it is used in this instance as a way to highlight the fact that building on other ideas is a central part of the OpenIDEO process, and as a way to incentivize attending to and describing the sources of one’s ideas. Further, 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 3**). This culture 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 my 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.

2.2 SAMPLE AND INITIAL DATA COLLECTION

The full dataset for this study consists of 2,341 concepts posted for 12 completed challenges by 1,190 unique contributors, citing 4,557 unique inspirations; 241 (10%) of these concepts are shortlisted for further refinement. These challenges were sampled for uniformity in platform features (OpenIDEO periodically adds/removes/refines features: e.g., citation features were added from the 6th challenge onwards; design phase structure was altered slightly after the 18th challenge onwards). See **Table 2** for a description of the 12 challenges (with some basic metadata on each challenge). **Figure 4** shows the full-text challenge brief for two challenges.

Table 2: Descriptions and number of posts for OpenIDEO challenges in final analysis sample

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

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 2,341 concepts, 707 (posted by 357 authors) cited at least one inspiration, collectively citing 2,245 unique inspirations. 110 of these concepts (~16%) were shortlisted (see Table 2 for a breakdown by challenge). This set of 707 concepts is the primary sample for this dissertation; 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. I analyze the impact of distance and diversity of *inspirations* (and not cited concepts) given my focus on ideation processes during "original" or non-routine design, where designers often start with a problem and only "inspirations" (information about the problem, potentially related designs) rather than routine design (e.g., configuration, 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). Also, the Conceptual Leap and Combination hypotheses map most clearly to non-routine design: the theoretical and research base is primarily concerned with the creation of new designs, rather than incremental modification or improvement of existing designs.

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 4: Full-text of challenge briefs from two OpenIDEO challenges.

2.3 MEASURES

2.3.1 Creative Outcome

The creative outcome measure is 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 considered by many authors to be a “gold standard” for measurement of innovation (Amabile, 1982; Baer & McKool, 2009; Brown, 1989; Sawyer, 2012). Further, the panel’s judgments combine consideration of both novelty and quality (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. I use shortlist (rather than win status) given my 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 and Diversity

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. 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. These issues are compounded when considering conceptual distance *between sources*, especially for concepts with many (e.g., more than two or three) sources, since all pairwise combinations need to be considered.

I address this methodological challenge by using probabilistic topic modeling (Blei, 2012; Steyvers & Griffiths, 2007), a major computational approach for understanding large collections of unstructured text. Topic modeling is 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 it emphasizes 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).

I 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. One can derive conceptual *similarity* between any pair of documents by computing the cosine between their topic-weight vectors. In essence, documents that have the same dominant topics in similar relative proportions are the most similar.

2.3.2.2 Document preprocessing

To train the topic model, I used all documents in the full dataset, i.e., 2,341 concepts, 4,557 inspirations, and 12 challenge briefs (6, 910 total documents). 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, I removed a standard list of stopwords, i.e., highly frequent words that do not carry semantic meaning on their own (e.g., “the”, “this”). I used the open-source MAchine Learning for LanguagE Toolkit’s (MALLET; McCallum, 2002) stopword list.

2.3.2.3 Model-building

I used MALLET to train an LDA model with 400 topics (LDA requires that the modeler pre-specify the number of topics to be learned), 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. Additional technical details on the model-building procedure are available in the Appendix A. Resulting cosines between inspirations and the challenge brief ranged from .01 to .91 (M = .21, SD = .18), a fairly typical range for large-scale information retrieval applications (Jessup & Martin, 2001).

2.3.2.4 Model Validation

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

Continuous similarity. Trained raters used a Likert-type scale to rate inspirations from two OpenIDEO challenges (bone-marrow and e-waste, n = 345 and 199, respectively) for similarity to their challenge brief, from 1 (very dissimilar) to 6 (extremely similar). I was able to train and obtain complete ratings from five raters for the e-waste challenge, and three for the bone-marrow challenge.

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").

Reflecting the difficulty of the task, raters achieved relatively low but acceptable agreement, with aggregate consistency intraclass correlation coefficient (ICC(2,3)) = .46 (mean inter-coder correlation = .26) for the bone-marrow challenge and ICC(2,5) = .74 (mean inter-coder correlation = .36) for the e-waste challenge. LDA cosines correlated highly with the continuous human similarity judgments for both challenges, with r = .54, 95% CI = [.46, .61] for

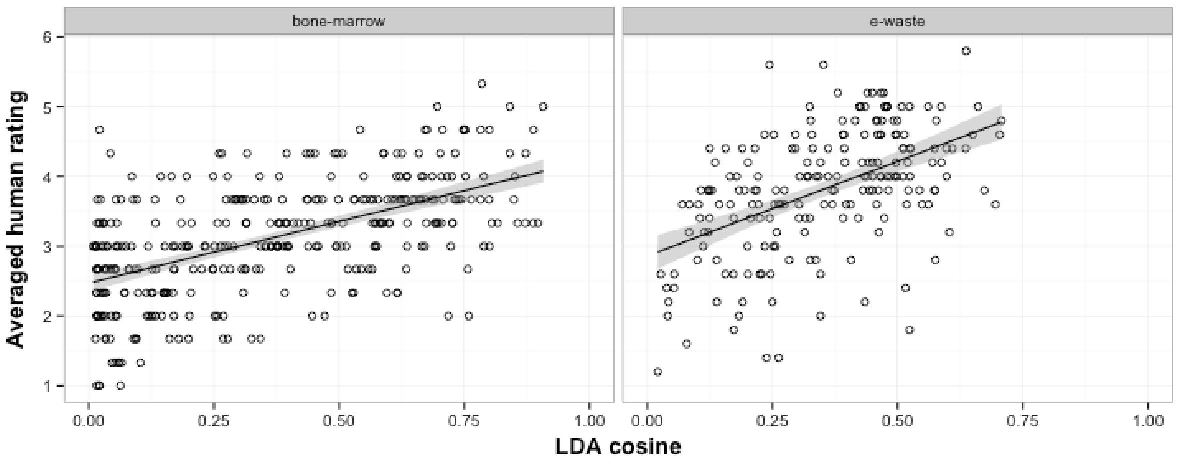

Figure 5. Scatterplot of LDA cosines vs. averaged human continuous similarity judgments for inspirations in the bone-marrow (left panel) and e-waste challenges (right panel).

the bone-marrow challenge, and r = .51, 95% CI = [.40, .60] for the e-waste challenge (see **Figure 5**). Note that in both challenges, the LDA-human correlation is better than the highest correlation between human raters (r = .39 for bone-marrow, and r = .48 for e-waste), reinforcing the value of automatic coding methods for this difficult task.

Binary distance. For comparability with prior work, I also measure fit with binary (within- vs. between-domain) distance ratings. Two raters also classified 345 inspirations from the same two challenges 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 .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

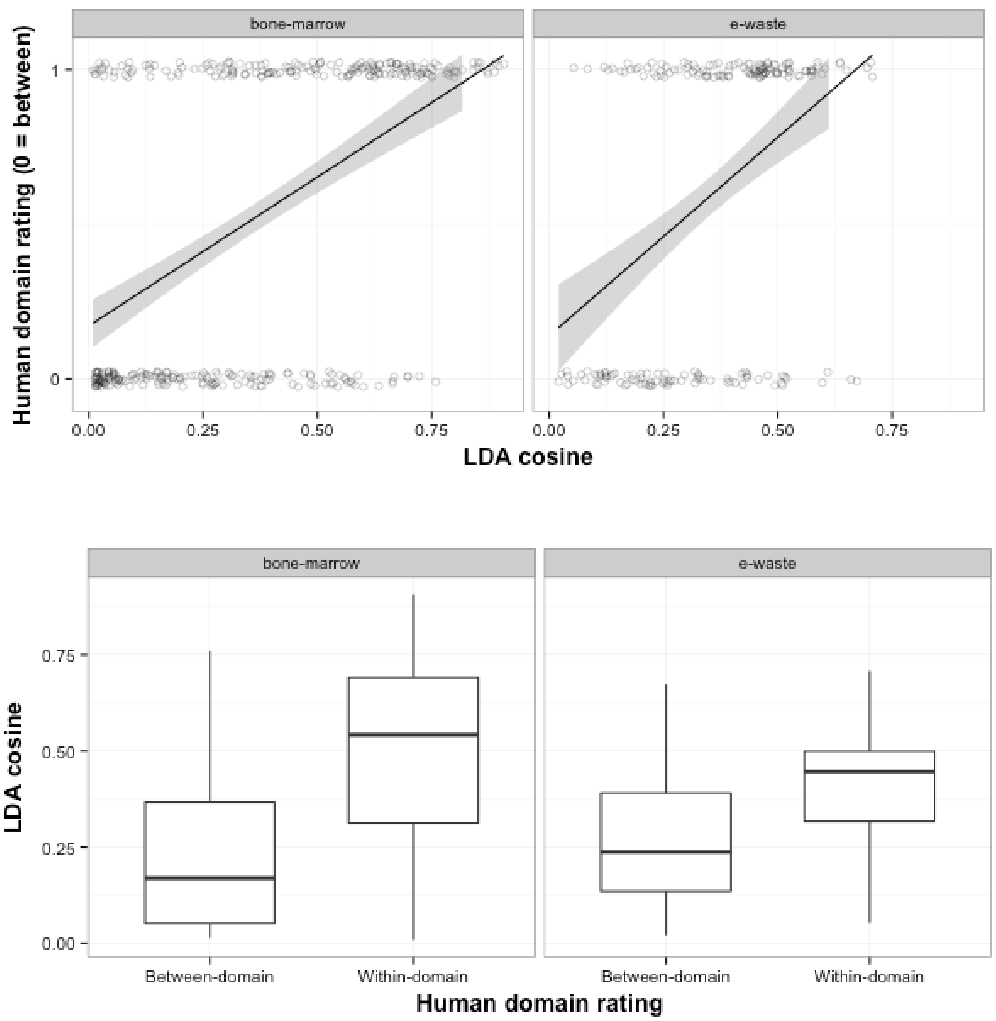

Figure 6. Scatterplot of LDA cosines vs. averaged human binary similarity judgments (top panel), and boxplot of cosine against the challenge brief for within- vs. between-domain inspirations (bottom panel).

high, at .50, 95% CI = [.42, .58] for the bone-marrow challenge, and .43, 95% CI = [.31, .54] for the e-waste challenge (see **Figure 6**, top panel). 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 6**, bottom panel).

Together, these results show that the LDA-derived cosines closely approximate human judgments of conceptual distance of inspirations from the challenge brief, and are therefore a reasonable substitute for those judgments.

Additional validation. As further validation, concepts within the same challenge were more similar to each other compared with concepts from a different challenge: the mean pairwise cosine for within-challenge pairs (M = 0.35, SD = 0.20 for first 14,700 pairwise comparisons

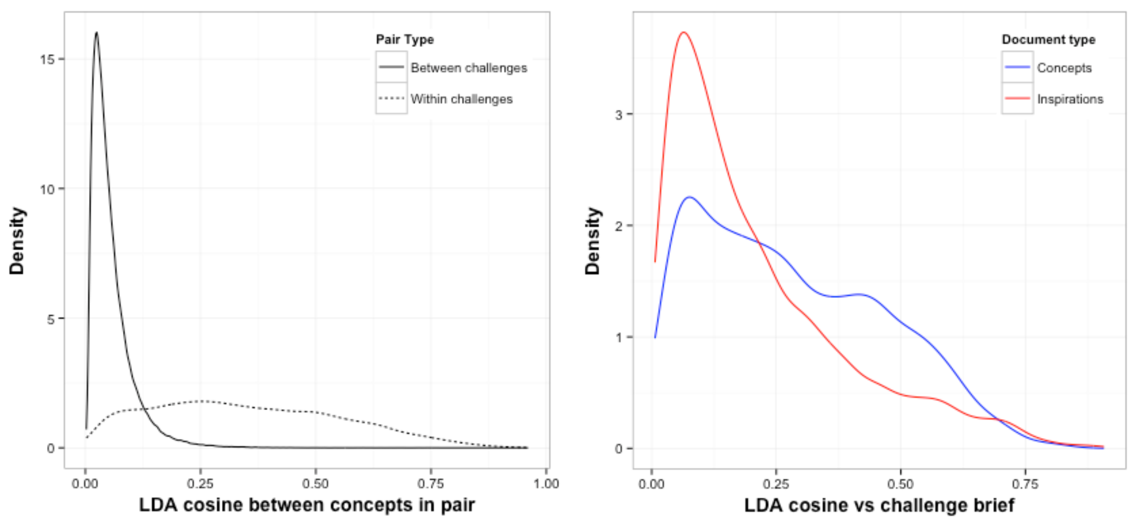

Figure 7. Gaussian kernel density plots for (A) pairwise cosines for between- and within-challenge concept pairs, and (B) cosines for concepts and inspirations vs. their challenge briefs.

between first 50 concepts) was much higher than that of between-challenge pairs (M = 0.06, SD = 0.05 for 333,000 pairwise comparisons with 550 concepts from remaining 11 challenges), Wilcoxon rank sum = 4,585,392,213, p < .0001, est. location difference in medians = 0.28 (see **Figure 7**A). Additionally, concepts were conceptually closer to the challenge brief (M = 0.28, SD = 0.19, N = 2340) compared to inspirations (M = 0.21, SD = 0.18, N = 4566), Wilcoxon rank sum = 6,609,964, p < .0001, est. location difference = 0.07 (see **Figure 7**B). This also validated the topic model because concepts are solutions to the problem, whereas inspirations may or may

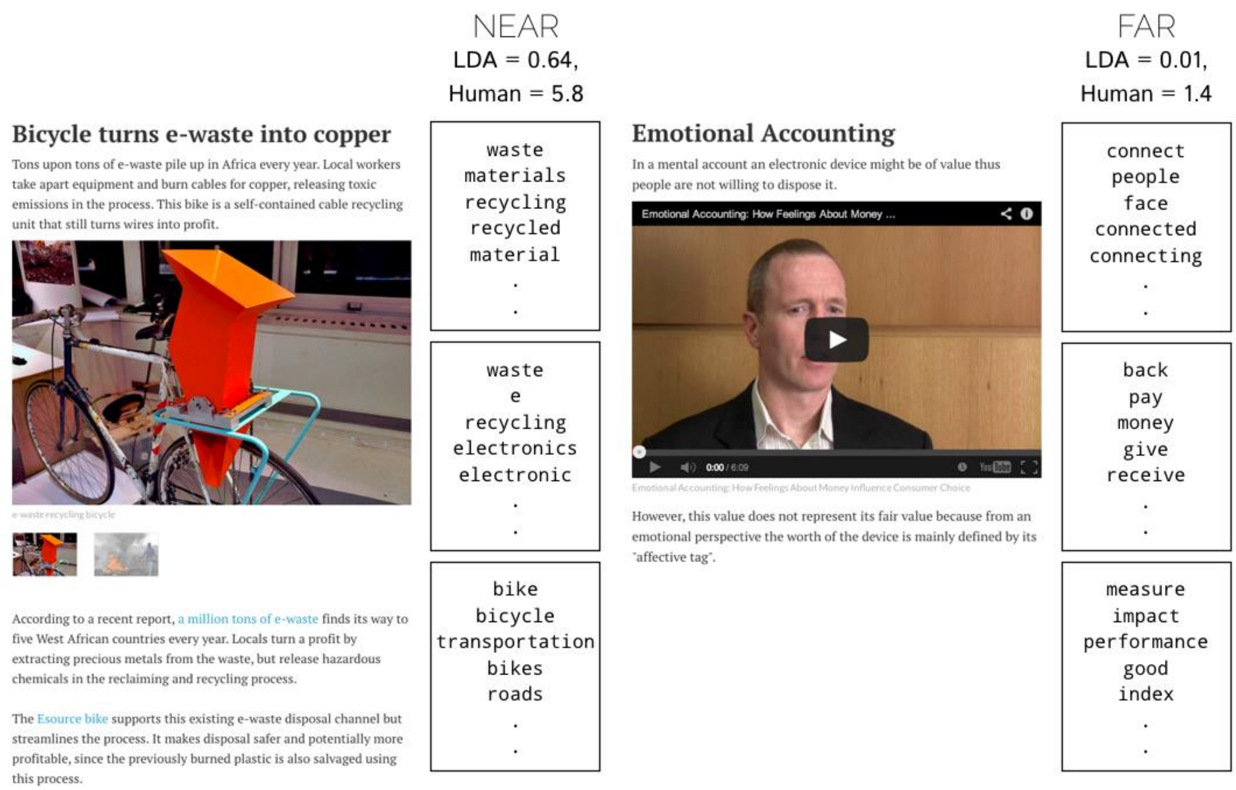

Figure 8. Topics found by LDA within examples of near and far inspirations for the e-waste challenge

not have closely related information; thus, concepts should overall be more similar to the challenge brief than inspirations. These results lend further strength to the validity of the topic model of this corpus.

Figure 8 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).

These LDA cosines are leveraged to generate the three primary measures of conceptual distance in subsequent chapters: distance from the problem, distance from one's solution path, and distance between sources (also called *diversity* of sources). The details of how these measures are derived will be given in context of those chapters.

2.3.3 Control measures

Given that the study design and analytic approach is primarily correlational, it is important to identify and rule out or adjust for major third variable factors that may influence the creative outcomes of concepts (particularly in the later stages, where prototyping and feedback are especially important) and may be correlated with the predictor variables.

Feedback. Considering the collaborative nature of OpenIDEO, I reasoned that feedback in the form of comments (labeled here as *FEEDBACK*) influence 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

The analytic goal is to predict the creative outcomes of 707 concepts, posted by 1,190 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 9**). 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 individual concepts[1]. Thus, I employ generalized linear mixed models (also called hierarchical or multilevel 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).

---

[1] Demographic variables were not available for all authors, and other author-level variables (e.g., number of contributions) were not predictive of mean Pr(shortlist) for authors.

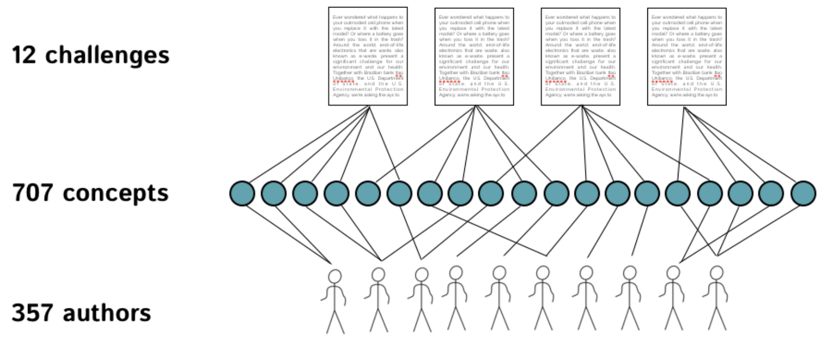

Figure 9. Illustrated cross-classified structure of the data

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 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)
- $u_{0authorj}$ and $u_{0challengek}$ model between-author and between-challenge variability in mean γ_{00}

I used the `lme4` package (Bates, Maechler, Bolker, & Walker, 2013) in R (R Core Team, 2013), using full maximum likelihood estimation by the Laplace approximation to fit the models. 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 ~.11 and .13, respectively, well above the cutoff recommended by Raudenbush and Bryk (2002)[2].

---

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

3.0 CONCEPTUAL DISTANCE AND CREATIVE SUCCESS

This chapter examines the Conceptual Leap Hypothesis (i.e., far sources provide the best insights for creative success).

3.1 METHODS

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 to 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, I 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.

I 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 examine possibly distinct effects of source distance: 1) *max* distance ($DIST_{MAX}$), i.e., the distance of the furthest source from the problem domain and 2) *mean* distance ($DIST_{MEAN}$). $DIST_{MAX}$ allows us to estimate "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 allows us to estimate the impact of the relative *balance* of relying on near vs. far sources (e.g., more near than far sources, or vice versa).

3.2 RESULTS

3.2.1 Descriptive statistics

On average, 16% of concepts in the sample get shortlisted (see **Table 3**). $DIST_{MEAN}$ 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 4**). All variables have significant bivariate correlations with *SHORTLIST* except for $DIST_{MAX}$; however, since it is a substantive variable of interest, and using bivariate correlations for feature selection can result in Type II error (e.g., predictors may have small but important effects that only become apparent after partialing out effects of other variables), I will model it nonetheless.

There do not appear to be potential multicollinearity concerns with the control variables, but there is a high bivariate correlation between $DIST_{MAX}$ and $DIST_{MEAN}$. The variance inflation

Table 3: Descriptive statistics for conceptual distance variables

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

Table 4: Bivariate correlations for conceptual distance variables

Variable	*FEEDBACK*	*SOURCE SHORT*	$DIST_{\mathrm{MAX}}$	$DIST_{\mathrm{MEAN}}$
SHORTLIST	0.33***	0.11**	–0.05	–0.10*
FEEDBACK		0.12**	0.07^{m}	0.02
SOURCESHORT			0.05	–0.05
$DIST_{\mathrm{MAX}}$				0.77***

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

factors for $DIST_{\text{MAX}}$ and $DIST_{\text{MEAN}}$ are 2.72 and 2.71, respectively (but drop to 1.02 and 1.02 when either is dropped), so I estimate separate models for the effects of $DIST_{\text{MAX}}$ and $DIST_{\text{MEAN}}$, each controlling for challenge- and author-nesting, *FEEDBACK*, and *SHORTSOURCE.*

3.2.2 Statistical models

I 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, I 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, $\chi^2(1) = 74.35$, $p = .00$.

3.2.2.1 Max distance

Adding $DIST_{\text{MAX}}$ to the model results in a significant reduction in deviance from the baseline model, $\chi^2(2) = 0.13$, $p = .47$ (see **Table 5**). This model estimated a *negative* effect of $DIST_{\text{MAX}}$, such that a 1-unit increase in $DIST_{\text{MAX}}$ predicted a .33 *decrease* in the log-odds of being shortlisted, after accounting for the effects of *FEEDBACK*, *SHORTSOURCE*, and challenge- and author-level nesting, $p < .05$. 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. Importantly, this negative effect of was robust across challenges: allowing $DIST_{MAX}$ to vary across challenges produced a near-zero estimate of

Table 5: 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) = $^{[lower, upper]}$

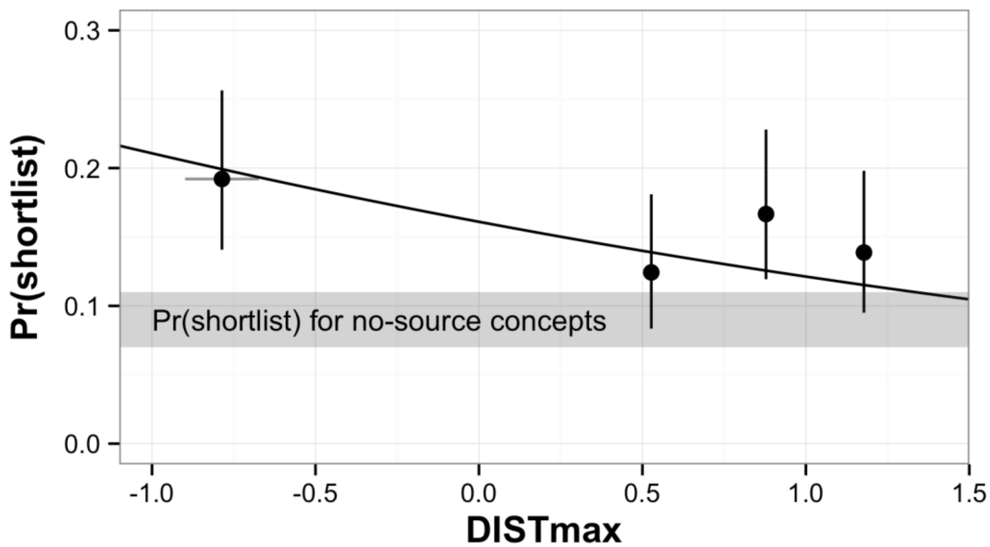

Figure 10. Model-fitted and observed relationship between $DIST_{MAX}$ and Pr(shortlist). Fitted values evaluated at mean values of feedback and source shortlist. Observed points are in equal N quartile bins. Vertical and horizontal error bars for points are 95% CI for Pr(shortlist) and $DIST_{MAX}$.

between-challenge variance in the effect of $DIST_{MAX}$; moreover, this model did not improve fit from the 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), and increased the Akaike Information Criterion (AIC).

Figure 10 visually displays the estimated relationship between $DIST_{MAX}$ and Pr(shortlist), evaluated at mean values of feedback and shortlisted sources. To aid interpretation, I 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 = .09, 95% CI = [.07 to .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 and observed curves with this bar highlights how the 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. I emphasize again that, despite the uncertainty in the degree of the negative relationship between $DIST_{\mathrm{MAX}}$ and Pr(shortlist), the data do *not* support an inference that the best ideas are coming from the farthest inspirations.

3.2.2.2 Mean distance

Similar results were obtained for $DIST_{\mathrm{MEAN}}$ (see **Table 6**). Adding $DIST_{MEAN}$ to the controls only model results in a small but significant reduction in deviance from the baseline model, $\chi^2(1)$ = 6.27, p = .01. There was a robust *negative* relationship between $DIST_{\mathrm{MEAN}}$ and Pr(shortlist), such that a 1-unit increase in $DIST_{\mathrm{MEAN}}$ was associated with a decrease of approximately .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 a 95% CI upper limit of at most B = –0.09 (but as high as –0.71). Again, as with $DIST_{\mathrm{MAX}}$, this negative relationship was robust and did not vary across challenges: allowing $DIST_{MEAN}$ to vary across challenges also produces a near-zero estimate of between-challenge variance in the effect of $DIST_{MEAN}$; similarly, AIC is increased with this model, and model fit does not improve, $\chi^2(2) = 0.07$, p = .48 (again, *p*-value here is halved to correct for overconservativeness).

Table 6: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on $\text{DIST}_{\text{MEAN}}$, 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]

As shown in **Figure 11**, as $DIST_{\text{MEAN}}$ increases, Pr(shortlist) approaches that of non-citing concepts, again suggesting (as with $DIST_{\text{MAX}}$) that the most beneficial sources appear to be ones that are relatively close to the challenge domain.

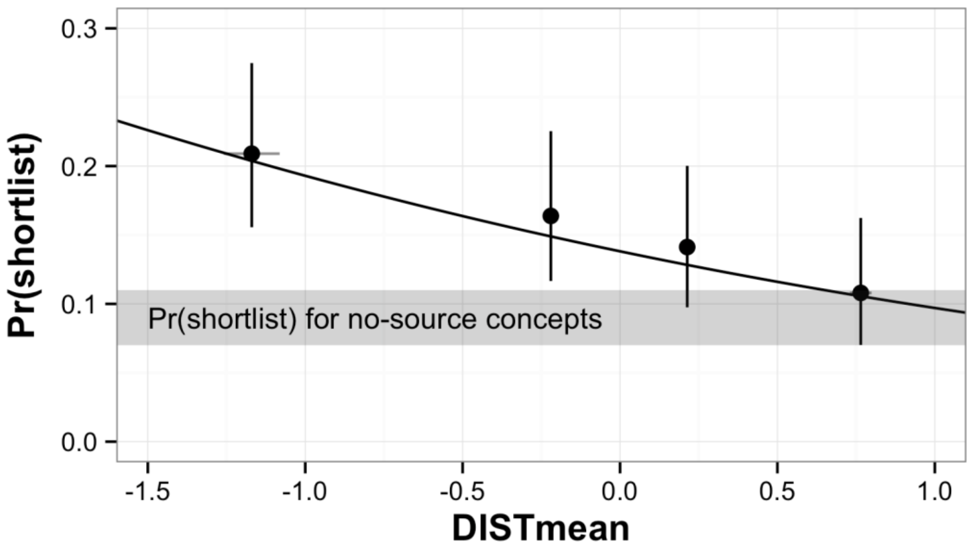

Figure 11. Model-fitted and observed relationship between $DIST_{MEAN}$ and Pr(shortlist). Fitted values evaluated at mean values of feedback and source shortlist. Observed points are in equal N quartile bins. Vertical and horizontal error bars for points are 95% CI for Pr(shortlist) and $DIST_{MEAN}$.

3.2.2.3 Robustness and sensitivity analysis

I first tested the robustness of these coefficient estimates to outliers 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), indicating that these results are unlikely to be unduly influenced by outlier concepts. Next, to address potential concerns about overfitting to singleton authors (i.e., authors with only one posted concept in the sample), I collapsed singleton authors into a single ID and refitted the fixed slope model. The resulting model was almost identical in terms of its fit to the data (deviance = 506.50) and coefficient estimate for $DIST_{MEAN}$ (B = –.39 [–.70, –.09]).

To address potential concerns about sensitivity to topic model parameter settings, I 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). Due to computational constraints, I checked robustness only for the models with $DIST_{MEAN}$. **Figure 12** shows the results of this analysis: attending first to the solid black dots (and their relationship to the red dashed line and the gray horizontal bar), we see that all models estimate

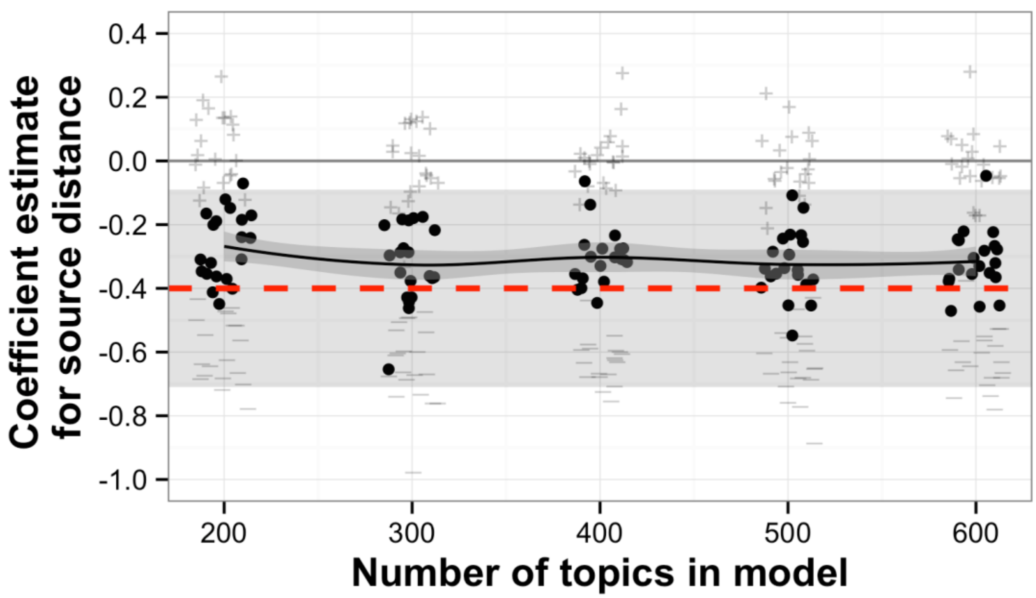

Figure 12. Coefficient estimate for $DIST_{mean}$ by topic model K. Solid black dots are point estimates for each model, with a loess line fitted to the relationship between K and point estimate size. Pluses and minuses are 95% lower and upper limits, respectively for those models. Dashed red line indicates point estimate for $DIST_{mean}$ reported above, with 95% CI represented with the horizontal gray bar.

negative coefficients for $DIST_{MEAN}$, with many of comparable magnitude to the model reported above, and almost all within the 95% confidence interval from the model above. Attending next to the loess smoothed line fitted to the point estimates and K, we see a relatively flat line, indicating that the sign and magnitude of the estimated effect of $DIST_{MEAN}$ do not appear to be dependent on K. Finally, attending to the spread of the pluses and minuses (relative to the horizontal gray bar), we see that the level of uncertainty in the magnitude of the effect is slightly larger when considering all 100 models, but is relatively independent of K. This robustness analysis thus shows that results reported here do not depend on a particular setting of K, and that the coefficient estimate, though slightly larger in magnitude than the larger sample of models, is not simply an outlier estimate, lending confidence in the robustness and validity of the results.

3.3 DISCUSSION

To summarize, the data provide no support for the Conceptual Leap Hypothesis; on the contrary, overreliance on far sources, measured by either $DIST_{\mathrm{MAX}}$ or $DIST_{\mathrm{MEAN}}$, is associated with worse innovative outcomes; said differently, the benefits of building on inspirations seem to accrue mainly for concepts that build more on near than far inspirations, with far inspirations that are not on the tail end of the distance continuum. Importantly, these effects were robust across challenges, addressing concerns raised about potential problem variation, at least among non-routine social innovation design problems. Additionally, addressing potential concerns about noise in my use of LDA to measure distance, a logistic regression model fitted with the two challenges for which I have human judgments of continuous distance (i.e., the bone-marrow and e-waste challenges) returns a very similar estimate of the effect of distance, albeit with more noise due to lack of statistical power (B = –0.65 [–2.13, 0.88]).

Some might be concerned with a lack of statistical power to detect problem variation, if it exists, given that I only have 12 challenges. The estimates of the size of the variance components for $DIST_{\mathrm{MAX}}$ and $DIST_{\mathrm{MEAN}}$ help to mitigate this concern, and plotting each coefficient by challenge (estimated with the random slope models; see **Figure 13**) shows that all challenges have negative coefficients, with very little variability between challenges. This gives us further confidence that the negative effects observed for $DIST_{\mathrm{MAX}}$ and $DIST_{\mathrm{MEAN}}$ are not driven by any particular challenge, but rather are consistent patterns observed across challenges.

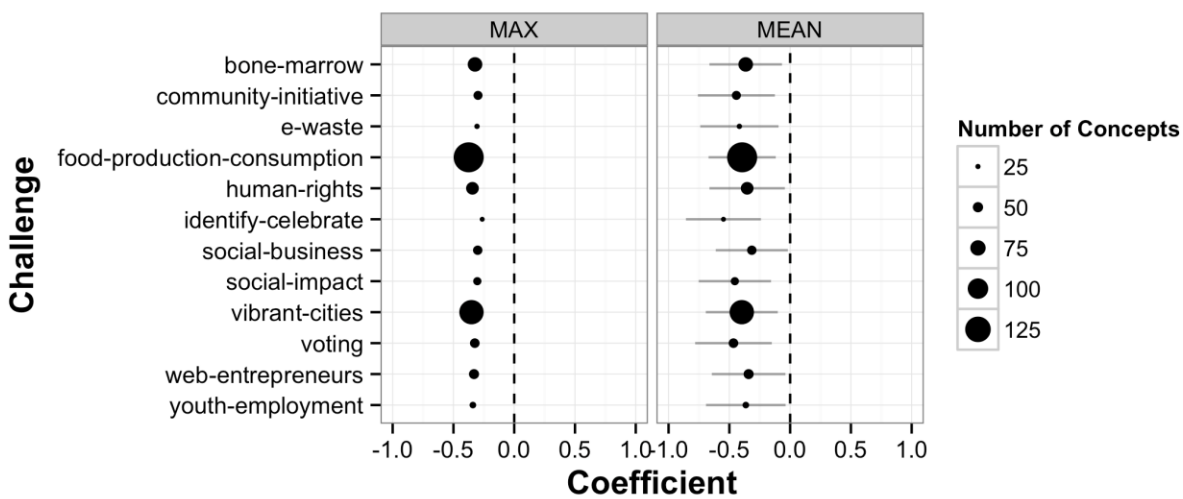

Figure 13. Coefficients for $DIST_{\mathrm{MAX}}$ and $DIST_{\mathrm{MEAN}}$ estimated by challenge with random slopes models, with 95% CIs. Dot size corresponds to the number of sampled concepts in the challenge, to give a sense of the challenge's contribution to the overall estimate.

4.0 CONCEPTUAL DISTANCE FROM SOLUTION PATH

This chapter considers an alternative conceptualization of source distance, i.e., defined with respect to the particular *solution* (or concept) one is considering. It could be argued that the breakthrough effect of far sources of inspiration may be most evident if the sources are far from one's "specific location" in the conceptual space (i.e., the specific concept being considered), rather than from one's "general location" (i.e., the general problem domain, as in Chapter 3). From a psychological standpoint, considering sources that are conceptually very different from the solution one is considering could cause one to reconsider one's solution approach, and explore novel iterations on (or alternatives to) one's solution approach; it is possible that sources that are far from the problem domain may nevertheless be familiar, whereas sources that are far from the particular concept may be more likely to be novel and inspiring. Conversely, one could be inspired by sources that are near to the problem but are nevertheless far from one's solution path (i.e., very different ideas than what one has previously considered, but are nevertheless still within the problem domain). Thus, problem distance might be too coarse a measure to capture the benefits of conceptual distance of sources, and analyzing how the distance of inspiration sources from their inspired *concept* (rather than the problem) may provide a more precise/sensitive test of the Conceptual Leap Hypothesis.

4.1 METHODS

Distance from self (hereafter denoted *DISTSELF*) was measured for each concept by measuring and reversing the cosine (i.e., subtracting from 0, to derive distance rather than similarity) between that concept and each of its cited inspirations. For analysis, these distances were summarized into two measures: 1) $DISTSELF_{MEAN}$, which is the mean of the distances, and 2) $DISTSELF_{MAX}$, which is the maximum of the distances (measuring the *furthest* a concept went from its own conceptual space).

Having established (in Chapter 3) distance from the problem as a useful predictor of Pr(shortlist), I now ask whether adding $DISTSELF_{MEAN}$ or $DISTSELF_{MAX}$ (I fit separate models because of their high intercorrelation) to the model improves our predictive power. I select $DIST_{MEAN}$ due to its slightly superior precision and fit.

4.2 RESULTS

4.2.1 Descriptive Statistics

Table 7 shows descriptive statistics for the *DISTSELF* predictors, along with their bivariate correlations with the other variables. Notably, they do not correlate strongly with $DIST_{MEAN}$, validating our choice to examine them separately rather than treating them as the same construct. No other strong correlations with the other predictors give initial cause for concern over

Table 7: Descriptive statistics for DISTSELF measures and correlations with other variables

	$DISTSELF_{MEAN}$	$DISTSELF_{MAX}$
Descriptives		
Min	-0.93	-0.93
Max	-0.01	-0.01
Median	-0.22	-0.11
Mean	-0.26	-0.18
SD	0.17	0.18
Correlations		
SHORTLIST	-0.06^{m}	-0.04
FEEDBACK	-0.01	0.05
SOURCESHORT	0.11**	0.13***
$DIST_{MEAN}$	0.13***	0.10**
$DISTSELF_{MEAN}$		0.86***

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

multicollinearity. A simple test of multicollinearity using a simple linear regression of *SHORTLIST* on all other covariates and $DISTSELF_{MAX}$ and $DISTSELF_{MEAN}$ separately confirms that multicollinearity is not a concern (variance inflation factor = 1.05 for both $DISTSELF_{MAX}$ and $DISTSELF_{MEAN}$). As with distance from the problem, $DISTSELF_{MAX}$ and $DISTSELF_{MEAN}$ are highly correlated, and estimated variance inflation factors of 4.15 and 4.13 motivate examining them separately.

4.2.2 Statistical Models

As before, I fit a series of generalized linear mixed models with `glmer` in R using full maximum likelihood estimation by the Laplace approximation, with concepts cross-classified within both authors and challenges. Both *DISTSELF* predictors were rescaled to range from –10 to 0 (by multiplying them by 10), since a 1-unit change on the original -1 to 0 scale would not be meaningful (i.e., would span the whole range of the variable).

4.2.2.1 Max distance from self

Consider first the model for $DISTSELF_{MAX}$. Adding $DISTSELF_{MAX}$ to the best-fitting model (with only controls and $DIST_{MEAN}$) from before results in a small reduction in deviance that also fails a likelihood ratio test of statistical significance, χ^2 (1) = 2.21, p = 0.14, and an increase in the AIC to 517.50 (see **Table 8**). The model estimates a very similar (albeit smaller and less precise) effect to $DIST_{MEAN}$, i.e., a slightly *negative* effect, with an increase of .10 in $DISTSELF_{MAX}$ associated with a decrease of approximately .19 in the log-odds of being shortlisted. **Figure 14** shows the best-fitting line relating $DISTSELF_{MAX}$ to Pr(shortlist), holding all other covariates at their mean values. Adding a random effect of challenge on $DISTSELF_{MAX}$ does not meaningfully decrease deviance from the simpler fixed effects model, χ^2 (2) = 0.08, p = .48 (p-value is halved to correct for overconservativeness), and also further increases AIC to 520.83.

Table 8: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on *DISTSELF*$_{\mathrm{MAX}}$, with comparison to baseline model (fixed *DIST*$_{MEAN}$)

	Baseline model (fixed $DIST_{MEAN}$)	$DISTSELF_{MAX}$, fixed slope	$DISTSELF_{MAX}$, random slope
Fixed effects			
γ_{00}, intercept	–2.74 [–3.36, –2.11]	–2.95 [–3.64, –2.26]	–2.95 [–3.91, –2.17]
γ_{10}, *FEEDBACK*	0.10*** [0.07, 0.12]	0.10*** [0.07, 0.12]	0.10*** [0.07, 0.12]
γ_{20}, *SOURCESHORT*	0.13 [–0.09, 0.35]	0.15 [–0.08, 0.37]	0.15 [–0.08, 0.38]
γ_{30}, $DIST_{MEAN}$	–0.40* [–0.71, –0.09]	–0.36* [–0.67, –0.05]	–0.36* [–0.69, –0.06]
γ_{40}, $DISTSELF_{MAX}$		–0.10 [–0.24, 0.03]	–0.10 [–0.27, 0.08]
Random effects			
$u_{0authorj}$ for intercept	0.31	0.30	0.29
$u_{0challengek}$ for intercept	0.73	0.74	0.80
$u_{3challengek}$ for $DISTSELF_{MAX}$			0.00
Model fit statistics			
Deviance	505.13	502.92	502.83
AIC	517.13	516.92	520.83

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

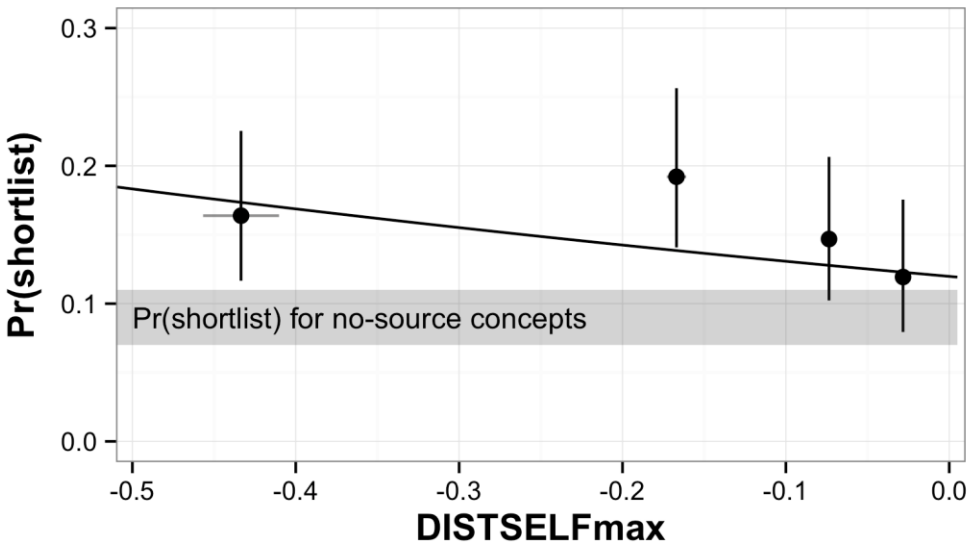

Figure 14. Model-fitted and observed relationship between $DISTSELF_{\mathrm{MAX}}$ and Pr(shortlist). Fitted values evaluated at mean values of feedback, source shortlist, and $DIST_{\mathrm{MEAN}}$. Observed points are in equal N quartile bins. Vertical and horizontal error bars for points are 95% CI for Pr(shortlist) and $DISTSELF_{\mathrm{MAX}}$.

4.2.2.2 Mean distance from self

Considering now $DISTSELF_{MEAN}$, the results are very similar to the analysis of $DISTSELF_{MAX}$. Adding $DISTSELF_{MEAN}$ to the best-fitting model from before results in a small reduction in deviance: however, this reduction fails a likelihood ratio test of statistical significance at the conventional .05 level, χ^2 (1) = 1.58, p = 0.21, and increases AIC from the previous best-fitting model's AIC of 517.13 (see **Table 9**). The lack of improvement notwithstanding, the model estimates a slightly *negative* effect of $DISTSELF_{MEAN}$, with an increase of .10 being associated with a decrease of approximately .10 in the log-odds of being shortlisted (note, however, that the confidence interval indicates that the effect could be very slightly positive). **Figure 15** shows the

Table 9: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on *DISTSELF*$_{MAX}$, with comparison to baseline model (fixed *DIST*$_{MEAN}$)

	Baseline model (fixed *DIST*$_{MEAN}$)	*DISTSELF*$_{MEAN}$, fixed slope	*DISTSELF*$_{MEAN}$, random slope
Fixed effects			
γ_{00}, intercept	–2.74 [–3.36, –2.11]	–3.02 [–3.78, –2.25]	–3.04 [–3.91, –2.17]
γ_{10}, *FEEDBACK*	0.10*** [0.07, 0.12]	0.10 [0.07, 0.12]	0.10 [0.07, 0.12]
γ_{20}, *SOURCESHORT*	0.13 [–0.09, 0.35]	0.14 [–0.08, 0.37]	0.14 [–0.08, 0.38]
γ_{30}, *DIST*$_{MEAN}$	–0.40* [–0.71, –0.09]	–0.36* [–0.67, –0.05]	–0.36* [–0.69, –0.06]
γ_{40}, *DISTSELF*$_{MEAN}$		–0.10 [–0.25, 0.05]	–0.10 [–0.27, 0.08]
Random effects			
$u_{0authorj}$ for intercept	0.31	0.30	0.30
$u_{0challengek}$ for intercept	0.73	0.73	1.24
$u_{3challengek}$ for *DISTSELF*$_{MEAN}$			0.02
Model fit statistics			
Deviance	505.13	503.55	502.37
AIC	517.13	517.55	520.37

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

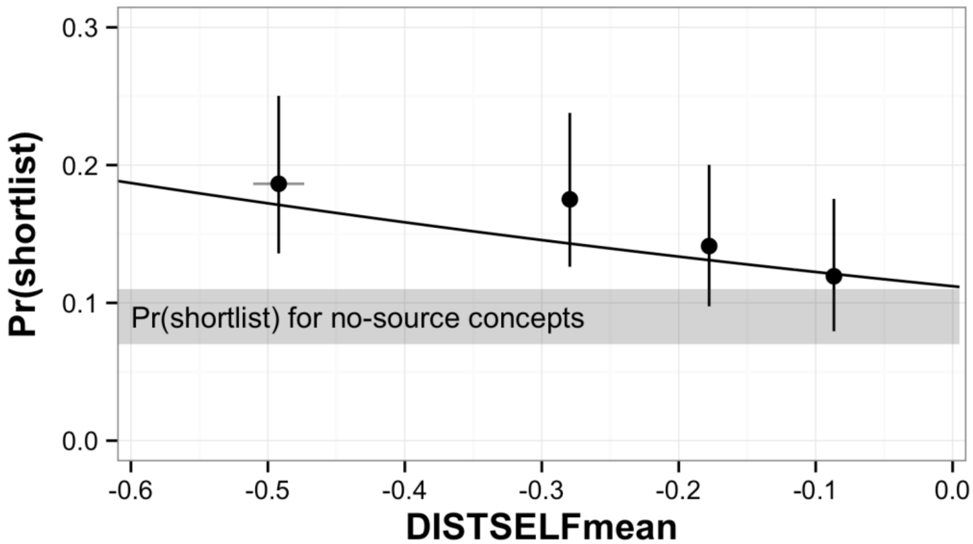

Figure 15. Model-fitted and observed relationship between $DISTSELF_{\mathrm{MEAN}}$ and Pr(shortlist). Fitted values evaluated at mean values of feedback, source shortlist, and $DIST_{\mathrm{MEAN}}$. Observed points are in equal N quartile bins. Vertical and horizontal error bars for points are 95% CI for Pr(shortlist) and $DISTSELF_{\mathrm{MEAN}}$.

best-fitting line relating $DISTSELF_{MEAN}$ to Pr(shortlist), holding all other covariates at their mean values.

As with $DIST_{MEAN}$, there does not appear to be problem variation: adding a random effect of challenge on $DISTSELF_{MEAN}$ does not meaningfully decrease deviance from the simpler fixed effects model, χ^2 (2) = 1.18, p = .28 (p-value is halved as before), and also further increases AIC to 520.37. The estimated challenge-level variance is also near-zero.

4.3 DISCUSSION

The analyses in this chapter were conducted to explore a potentially more precise test of the Conceptual Leap Hypothesis, with the assumption that problem distance might be too coarse a measure to capture the true benefits of conceptual distance (which might accrue from soruces that are far from one's solution path, not the problem *per se*).

Overall, the results of this analysis were very similar to that of the previous chapter. Despite the relative statistical independence of the two kinds of distance measures (problem vs. self, with $r < .15$), we saw very similar (albeit significantly smaller and noisier) trends in the negative direction for the effect of the *DISTSELF* measures. Thus, as before, I find no support for the Conceptual Leap Hypothesis, and instead find a similar opposition to it, i.e., greater distance of sources appears to *decrease* creative success. No significant problem variation was detected. It is worth noting that the effect of $DIST_{MEAN}$, remained robust to the inclusion of the *DISTSELF* measures in the models.

These results, then, both strengthen and broaden the findings from Chapter 3. There remains strong evidence that building more on sources that are far from the *problem* is associated with lower creative success, and that this is not simply an artifact of a noisy measure that is poorly calibrated to theory: even when we consider distance directly from one's solution path, overreliance on far sources can harm creative success.

5.0 CONCEPTUAL DIVERSITY

The previous two chapters examined the Conceptual Leap Hypothesis from two complementary angles, measuring distance from both the problem and self. This chapter examines the Conceptual Combination Hypothesis (i.e., far combinations of sources provide better insights for creative breakthroughs than near combinations), and also seeks to further examine the robustness of the negative effect of problem distance, and whether or how it might interact with conceptual diversity of sources.

5.1 METHODS

5.1.1 Sample

The sample for this analysis is a subset of the 707 concepts that cite at least 2 inspirations (since diversity is undefined for sets of size < 2). The 456 concepts that cite at least 2 inspirations constitute the sample for this analysis (see **Table 10** for a breakdown by challenge). It is important to note that statistical power is now likely to be severely reduced, not simply because of the overall reduction in N (and also by challenge), but also the difficulty of estimating predictions for so few shortlist cases (< 4 for four challenges).

Table 10: Descriptions and number of posts for OpenIDEO challenges in diversity analysis sample

Challenge	Num. inspirations *(% diff from previous)*	Num. concepts *(% diff from previous)*	Num. shortlisted *(% diff from previous)*
Bone-marrow	170 *(9%)*	31 *(56%)*	3 *(57%)*
Community-initiative	159 *(1%)*	36 *(18%)*	9 *(18%)*
E-waste	58 *(3%)*	18 *(31%)*	5 *(38%)*
Food-production-consumption	256 *(4%)*	85 *(42%)*	7 *(30%)*
Human-rights	246 *(1%)*	45 *(27%)*	6 *(14%)*
Identify-celebrate	119 *(2%)*	14 *(42%)*	8 *(38%)*
Social-business	126 *(4%)*	38 *(17%)*	8 *(27%)*
Social-impact	63 *(6%)*	24 *(40%)*	7 *(42%)*
Vibrant-cities	546 *(2%)*	81 *(32%)*	11 *(15%)*
Voting	236 *(2%)*	32 *(32%)*	4 *(50%)*
Web-entrepreneurs	76 *(14%)*	28 *(43%)*	2 *(71%)*
Youth-employment	112 *(5%)*	24 *(25%)*	2 *(33%)*

5.1.2 Measures

Diversity (hereafter denoted *DIV*) was measured for each concept by measuring and reversing all pairwise cosines (i.e., subtracting from 0, to derive distance rather than similarity) between inspirations cited by that concept.

5.2 RESULTS

5.2.1 Descriptive Statistics

Table 11 summarizes the descriptive statistics and intercorrelations between the variables. There are statistically significant positive correlations between the control variables and Pr(shortlist). There are no strong inter-correlations between the predictor variables, alleviating potential concerns about multicollinearity; a variance inflation analysis also shows that having *DIV* and $DIST_{\text{MEAN}}$ in the same model should not introduce multicollinearity, with variance inflation factors of 1.16 for both variables.

Table 11: Descriptive statistics and intercorrelations between diversity variables

	Descriptives	Correlations			
Variable	M *(SD)*	*FEEDBACK*	*SOURCE SHORT*	$DIST_{\text{MEAN}}$	*DIV*
SHORTLIST	0.16 *(0.36)*	0.33***	0.11**	–0.10*	–0.01
FEEDBACK	9.14 *(9.92)*		0.12**	0.02	0.05
SOURCESHORT	0.61 *(1.07)*			–0.05	0.10*
$DIST_{\text{MEAN}}$	–0.13 *(0.62)*				0.29***
DIV	2.02 *(1.25)*				–

5.2.2 Statistical Models

As before, I fit a series of generalized linear mixed models with `glmer` in R using full maximum likelihood estimation by the Laplace approximation, with concepts cross-classified within both authors and challenges. I rescale *DIV* (multiplying it by 10) for easier interpretation (a more meaningful "1-unit" change).

As before, I find that there is significant nesting for the reduced set of 456 concepts. The intraclass correlation coefficient (ICC) for author-nesting is approximately 0.14 (again using the approximation for level-1 residuals from Zeger et al, 1988), indicating that approximately 14% of the total variability in Pr(shortlist) lies between authors. Similarly, the ICC estimate for challenge-nesting is approximately 0.09, indicating that approximately 9% of the total variability in P(shortlist) lies between challenges. Both ICC values are well above conventional cut-offs for ICCs (e.g., as recommended by Raudenbush & Bryk, 2002), and both higher-level random effects are statistically significant using a nested likelihood ratio test (comparing the cross-classified with a challenge-nesting only and author-nesting only model), $\chi^2(1) = 4.41$, $p < .05$ and $\chi^2(1) = 4.52$, $p < .05$, for author- and challenge-level variance respectively. **Table 12** presents the model estimates and fit statistics for these models. As before, I use a model with the controls variables and $DIST_{MEAN}$ as a baseline for comparing what is added by *DIV*. The baseline model gives a large and statistically significant reduction in deviance compared to the null model, χ^2 (2) = 64.70, $p = 0.00$. Adding a fixed slope for *DIV* to this model does not provide any meaningful reduction in deviance, with the likelihood ratio being essentially zero,

Table 12: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on *DIV*, with comparison to baseline model (controls and $DIST_{MEAN}$)

	Baseline model (controls and $DIST_{MEAN}$)	With *DIV*, fixed slope	With *DIV* random slope
Fixed effects			
γ_{00}, intercept	–3.08 $^{[–3.37, –2.12]}$	–3.05 $^{[–3.99, –2.12]}$	–3.03 $^{[–4.11, –1.95]}$
γ_{10}, *FEEDBACK*	0.10*** $^{[0.07, 0.12]}$	0.10*** $^{[0.07, 0.13]}$	0.10*** $^{[0.07, 0.13]}$
γ_{20}, *SOURCESHORT*	0.25^{m} $^{[–0.10, 0.35]}$	0.25^{m} $^{[–0.03, 0.52]}$	0.26^{m} $^{[–0.03, 0.54]}$
γ_{30}, $DIST_{MEAN}$	$–0.49^{m}$ $^{[–0.71, 0.10]}$	$–0.50^{m}$ $^{[–1.05, 0.04]}$	–0.54* $^{[–1.08, –0.00]}$
γ_{40}, *DIV*		0.01 $^{[–0.27, 0.30]}$	0.03 $^{[–0.28, 0.33]}$
Random effects			
$u_{0authorj}$	0.47	0.47	0.44
$u_{0challengek}$	0.71	0.71	1.63
$u_{1challengek}$			0.05
Model fit statistics			
Deviance	323.57	323.57	321.74
AIC	335.57	337.57	339.74

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

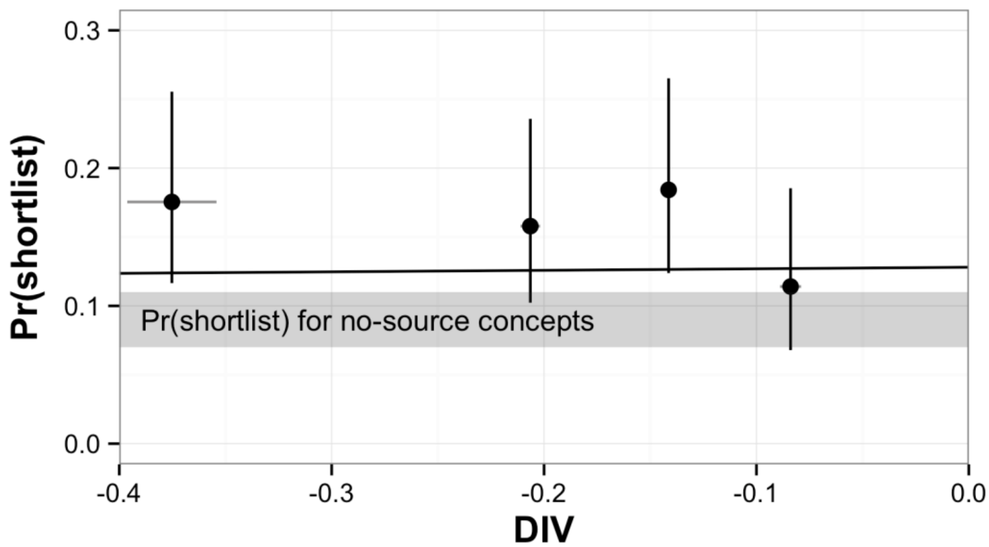

Figure 16. Model-fitted and observed relationship between *DIV* and Pr(shortlist). Fitted values evaluated at mean values of feedback, source shortlist, and $DIST_{MEAN}$. Observed points are in equal N quartile bins. Vertical and horizontal error bars for points are 95% CI for Pr(shortlist) and *DIV*.

χ^2 (1) = 0.00, p = 0.92, and an increase in the AIC. The point estimate for the effect of a change in .10 (remember that *DIV* is rescaled in this model) is also essentially zero (albeit with a fairly wide confidence interval). To ensure that this wide confidence interval is not due to large problem variation, I estimate an additional model with a random slope for *DIV*. This model estimates essentially zero challenge-variance, does not meaningfully decrease variance from the fixed slope model, χ^2 (2) = 1.83, p = .23 (p-value is halved, as before), and also further increases AIC. **Figure 16** shows this estimated zero effect of *DIV* in relation to observed Pr(shortlist) for 4 equal N bins (with 95% CIs). There does not appear to be a discernible trend, except perhaps a slight drop-off in Pr(shortlist) at extreme values of *DIV* (i.e., when most sources are very far from each other).

5.3 DISCUSSION

The purpose of this chapter was to test the Conceptual Combination Hypothesis, and also further explore the robustness of the negative effect of problem distance. Overall, this analysis did not support the Conceptual Combination Hypothesis, although the high attrition from the initial sample of 707 concepts, and the wide confidence intervals (relative to the estimated effect) give cause for caution in interpreting this as a “strong null” (i.e., confident estimated zero effect).

With regards to the robustness of the effect of problem distance, the estimated negative effect appears to be robust to conceptual diversity, although the uncertainty in the estimate is higher, probably due in part to the lack of statistical power. I also find no evidence that they interact in important ways, at least in this data, although the lack of statistical power means interactions should not necessarily be ruled out for future analyses. Thus, these results continue to strengthen the findings from Chapter 3: regardless of conceptual diversity, overreliance on sources that are conceptually far from one’s problem leads to lower creative success, contrary to the Conceptual Leap Hypothesis.

6.0 INDIRECT EFFECTS OF DISTANCE

In this chapter I consider the possibility of “indirect effects”. The analyses in this chapter are inspired in part by von Wartburg, Teichert, and Rost’s (2005) multi-stage analysis of patent citation paths and technological lineages. There are theoretical reasons to suppose that considering indirect sources may provide additional, potentially different, insights into the effects of the conceptual distance inspiration sources. For instance, it could be that the benefits of distance can often be overwhelmed by the cognitive costs of mapping/adapting far sources, or sifting through potentially irrelevant inferences from far sources. Concepts that build on other inspirations or concepts that have already mapped or processed these far sources may be able to benefit from their sparks without paying the costs of being the first to process them. Relatedly, far sources may not yield immediately usable ideas: they may be novel, but require additional processing in order to be useful for the problem. Thus, I might expect to see different results when considering the conceptual distance of *indirect* sources (i.e., sources cited by immediately cited sources), more in line with the Conceptual Leap Hypothesis. Let us first consider effects of distance from the problem.

6.1 METHODS

6.1.1 Measures

To gather indirect sources for a given concept, a conceptual genealogy for that concept was constructed via breadth-first search through the citation graph gathered in initial data collection: this search first returned all sources that concept built upon, and then returned all sources that each of these sources built upon (whether they were concepts or inspirations), traversing the conceptual tree to its endpoint. This search procedure was programmed to ignore duplicate entries: for instance, if an inspiration I was a direct source for a concept C (at level 1), and also for another concept/inspiration at level 2, it would only be counted once as a level 1 source for C.

I defined "indirect" inspirations as inspirations from levels 2 to 4 of each concept's genealogy (see **Figure 17**): this cut-off, while seemingly arbitrary, reflects our goal of examining the effects of sources that are "just recent enough" to have discernible effects (we may not be able to distinguish the effects of sources that are too deep in a genealogy), while having sufficient genealogical depth to allow for iteration and "preprocessing" of sources to occur. Notice from **Figure 17** that indirect sources would also include inspirations cited by cited *concepts* (i.e., the sources of concepts that acted as immediate sources for the root concept). One way to think about this relationship of the root concept with these indirect sources of other concepts is that (at least part of) the insights/information/ideas contained in those inspirations are "passed on" to the root concept through their incorporation into the concepts immediately cited by the root concept.

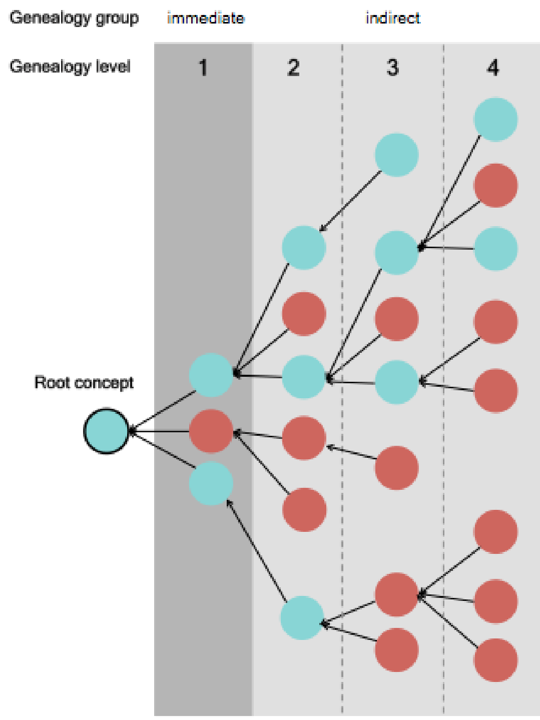

Figure 17. Illustrated example conceptual genealogy and operationalization of "indirect" sources as sources in levels 2 to 4 of the genealogy. Teal circles denote concepts; maroon circles denote inspirations.

As before, I computed *max* and *mean* distance measures for each inspiration source set (here denoted $IND\text{-}DIST_{MAX}$ and $IND\text{-}DIST_{MEAN}$.

6.1.2 Sample

Because I wanted to control for effects of immediate distance of sources, and not all immediately cited inspirations themselves cited inspirations, the sample for this analysis was reduced from 707 to 568 total concepts that *both* cited at least one immediate inspiration *and* included at least one inspiration in levels 2 to 4 of their genealogy. **Table 13** gives the breakdown of concepts

Table 13: Descriptions and number of posts for OpenIDEO challenges in indirect distance analysis sample

Challenge	Num. concepts *(% diff from previous)*	Num. shortlisted *(% diff from previous)*
Bone-marrow	54 *(24%)*	5 *(29%)*
Community-initiative	34 *(23%)*	7 *(36%)*
E-waste	23 *(12%)*	7 *(13%)*
Food-production-consumption	110 *(25%)*	8 *(20%)*
Human-rights	55 *(11%)*	6 *(14%)*
Identify-celebrate	13 *(46%)*	9 *(31%)*
Social-business	39 *(15%)*	9 *(18%)*
Social-impact	34 *(15%)*	8 *(33%)*
Vibrant-cities	104 *(13%)*	12 *(8%)*
Voting	40 *(15%)*	8 *(0%)*
Web-entrepreneurs	35 *(29%)*	4 *(43%)*
Youth-employment	27 *(16%)*	2 *(33%)*

(shortlisted and not shortlisted) by challenge, with notes on attrition levels from the initial sample of 707 concepts.

6.2 RESULTS

6.2.1 Descriptive Statistics

Descriptive statistics are shown in **Table 14**, and bivariate correlations in **Table 15**. Note that the overall mean Pr(shortlist) remains substantially similar to the initial sample of 707 concepts (i.e., ~15% compared to 16% in the original sample). No bivariate correlations give cause for concern over multicollinearity, and it is interesting to note that the *IND-DIST* measures are only weakly (if at all) related to $DIST_{MEAN}$, giving some confidence that they measure a (at least statistically) distinct construct.

Table 14: Descriptive statistics for indirect distance variables

Variable	Valid N	Min	Max	Mean	Median	SD
SHORTLIST	568	0	1	0.15	0	0.36
FEEDBACK	568	0	67	8.78	6	9.81
SOURCESHORT	568	0	11	0.63	0	1.04
$DIST_{MEAN}$	568	-2.93	1.67	-0.11	-0.01	0.74
$IND\text{-}DIST_{MAX}$	568	-2.65	1.90	0.92	1.02	0.56
$IND\text{-}DIST_{MEAN}$	568	-2.65	1.26	-0.11	-0.04	0.43

Table 15: Intercorrelations between indirect distance variables

Variable	*FEEDBACK*	*SOURCE SHORT*	$DIST_{\text{MEAN}}$	$IND\text{-}DIST_{\text{MAX}}$	$IND\text{-}DIST_{\text{MEAN}}$
SHORTLIST	0.33***	0.13**	–0.12**	0.00	–0.05
FEEDBACK		0.12**	–0.01	0.03	0.02
SOURCESHORT			–0.06	0.21***	0.01
$DIST_{\text{MEAN}}$				–0.04	0.10*
$IND\text{-}DIST_{\text{MAX}}$					0.54***

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

6.2.2 Statistical Models

As before, I use a model with the control variables and $DIST_{MEAN}$ as a baseline for comparing what is added by *IND-DIV*. This baseline model gives a large and statistically significant reduction in deviance compared to the null model, χ^2 (3) = 65.77, p = 0.00, and the estimated effects of all predictors are substantially similar to those estimated with the full sample.

6.2.2.1 Max

Adding a fixed slope for $IND\text{-}DIST_{MAX}$ to this model does not meaningfully reduce deviance, χ^2

Table 16: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on IND-DISTmax, with comparison to baseline model (controls and DISTmean)

	Baseline model (controls and *DIST*$_{MEAN}$)	Model 3: *IND-DIST*$_{MAX}$, fixed slope	Model 4: *IND-DIST*$_{MAX}$ random slope
Fixed effects			
γ_{00}, intercept	–2.78 [–3.43, –2.11]	–2.65 [–3.42, –1.89]	–2.68 [–3.42, –1.89]
γ_{10}, *FEEDBACK*	0.09*** [0.06, 0.12]	0.09*** [0.06, 0.12]	0.09*** [0.06, 0.12]
γ_{20}, *SOURCESHORT*	0.16 [–0.08, 0.39]	0.17 [–0.07 0.42]	0.18 [–0.07 0.42]
γ_{30}, *DIST*$_{MEAN}$	–0.42* [–0.78, –0.06]	–0.42* [–0.78, –0.06]	–0.42* [–0.78, –0.06]
γ_{40}, *IND-DIST*$_{MAX}$		–0.14 [–0.63, 0.35]	–0.13 [–0.63, 0.35]
Random effects			
$u_{0authorj}$	0.13	0.13	0.14
$u_{0challengek}$	0.72	0.71	0.47
$u_{1challengek}$			0.05
Model fit statistics			
Deviance	400.53	400.23	399.74
AIC	412.53	414.23	417.74

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

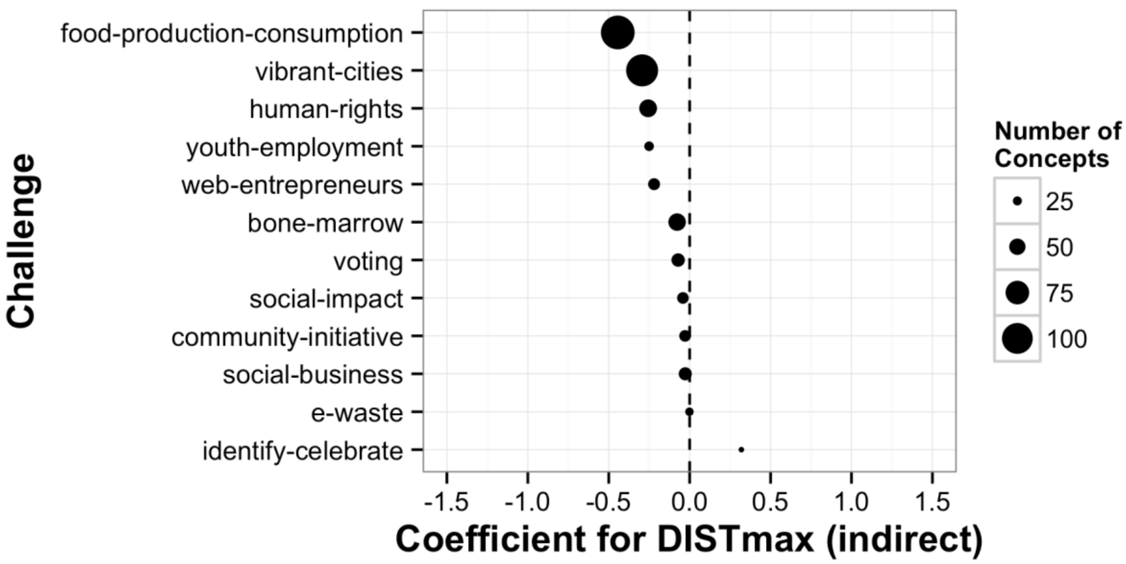

Figure 18. Coefficients for IND-DISTmax estimated by challenge with random slopes models, with 95% CIs, with dots sized by number of sampled concepts in challenge.

(1) = 0.31, p = 0.58, and results in a slight negative estimate for the effect of $IND\text{-}DIST_{MAX}$ (see **Table 16**); however, there is high uncertainty about this estimate, as indicated by the wide confidence interval. To ensure that this high uncertainty is not due to problem variation, I also estimated a random effect of challenge on the slope of $IND\text{-}DIST_{MAX}$. This model estimates a near-zero challenge-level variance component, and does not meaningfully reduce deviance from the fixed slope model, χ^2 (2) = 0.49, p = 0.39 (halved).

Figure 18 shows that the estimated effects of $IND\text{-}DIST_{MAX}$ for each challenge are relatively uniform, with most being either near-zero or slightly negative (and only one estimated slightly positive effect).

6.2.2.2 Mean

Different results were found for *IND-DIST*$_{MEAN}$ (see **Table 17**). While adding a fixed slope for

Table 17: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on *IND-DIST*$_{MEAN}$, with comparison to baseline model (controls and *DIST*$_{MEAN}$)

	Baseline model (controls and *DIST*$_{MEAN}$)	Model 3: *IND-DIST*$_{MEAN}$, fixed slope	Model 4: *IND-DIST*$_{MEAN}$ random slope
Fixed effects			
γ_{00}, intercept	–2.78 [–3.43, –2.11]	–2.65 [–3.42, –1.89]	–2.87 [–3.42, –1.89]
γ_{10}, *FEEDBACK*	0.09*** [0.06, 0.12]	0.09*** [0.06, 0.12]	0.09*** [0.06, 0.12]
γ_{20}, *SOURCESHORT*	0.16 [–0.08, 0.39]	0.17 [–0.07 0.42]	0.18 [–0.07 0.42]
γ_{30}, *DIST*$_{MEAN}$	–0.42* [–0.78, –0.06]	–0.42* [–0.78, –0.06]	–0.44* [–0.78, –0.06]
γ_{40}, *IND-DIST*$_{MEAN}$		–0.14 [–0.63, 0.35]	–0.21 [–0.63, 0.35]
Random effects			
$u_{0authorj}$	0.13	0.13	0.15
$u_{0challengek}$	0.72	0.71	0.97
$u_{1challengek}$			0.60
Model fit statistics			
Deviance	400.53	400.23	395.27
AIC	412.53	414.23	413.27

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

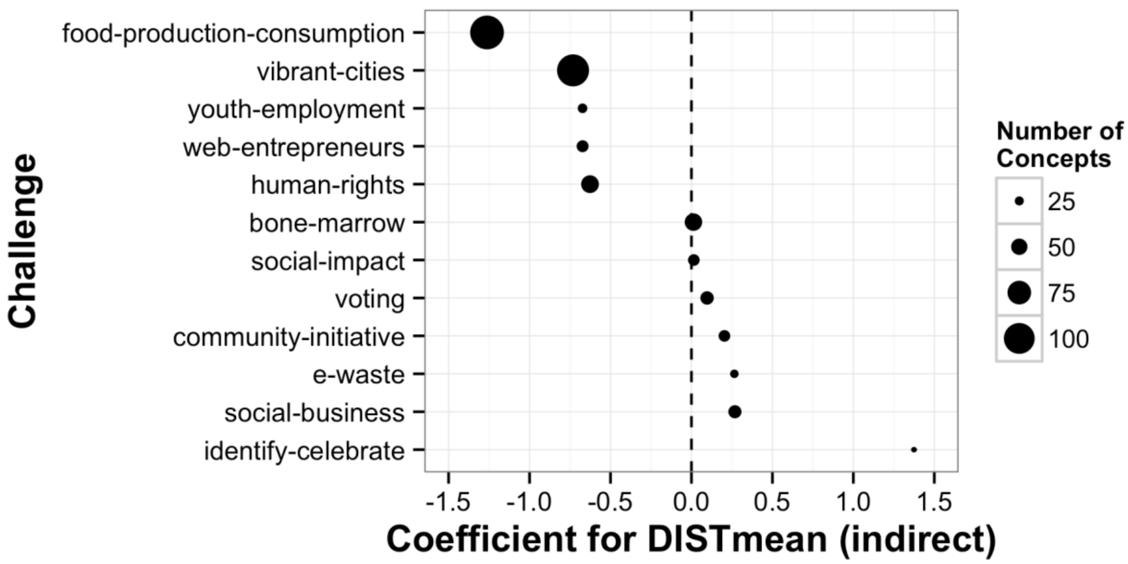

Figure 19. Coefficients for *IND-DIST*$_{\text{MEAN}}$ estimated by challenge with random slopes models, with 95% CIs, with dots sized by number of sampled concepts in challenge.

IND-DIST$_{MEAN}$ did not meaningfully reduce deviance, χ^2 (1) = 0.31, p = 0.58, fitting an additional model with a random effect of challenge on the slope of *IND-DIST*$_{MEAN}$ *did* meaningfully reduce deviance from the fixed slope model, χ^2 (2) = 4.64, p = 0.049 (halved), and estimated a challenge-level variance component of comparable size to the challenge-level variance in Pr(shortlist). However, this random-slopes model's reduction in deviance from the baseline model falls short of conventional statistical significance, χ^2 (3) = 5.27, p = 0.15, and has a slightly higher AIC (413.27 vs. 412.53).

Figure 19 shows that there appears to be a relatively even split between the challenges in terms of a positive vs. negative effect of *IND-DIST*$_{MEAN}$: five challenges have a negative coefficient (in line with the analyses of immediate sources; note, however, the increased size of the coefficients), five challenge have a relatively small positive coefficient, and two have a near-zero coefficient.

6.3 DISCUSSION

This chapter examined the possibility that the effects of problem distance might be different for *indirect* sources (e.g., additional iteration might be necessary to "convert" novel but raw ideas inspired by far sources into good ideas). The analyses showed that the effects of indirect *max* distance look very similar to that of immediate max distance (negative effect with no problem variation), albeit with far greater uncertainty (and a smaller coefficient). In contrast, there are some hints of problem-variant effects of indirect *mean* distance: at least for *some* challenges, there seems to be an estimated *positive* effect of $IND\text{-}DIST_{MEAN}$, more in line with the Conceptual Leap Hypothesis, and in keeping with the idea that the benefits of far sources might only start to show if they have been "preprocessed" by other ideas before being built upon.

What could explain this problem-variation? One conservative interpretation would be that the variation (specifically, the estimated positive effects) is an artifact of imprecise estimates for the smaller challenges (with fewer observations) in this sample: indeed, as is evident in **Figure 19**, the challenges with larger N do tend to have a more negative coefficient, $r = –.71$ [–.91, –.23], $p < .01$. In light of the robust negative effect of immediate distance of sources, one might suspect that the more positive estimates from the smaller challenges might be statistical flukes, and be more inclined to trust the negative estimates from the larger challenges: by analogy to statistical power and precision, with fewer samples of ideas built on far indirect-sources, there is a higher chance that the mean effect of those sources appears to be positive, but perhaps we begin to converge on the true distribution (i.e., mean negative effect) with more samples.

Alternatively, it may be more difficult for the expert panels to find "hidden gems" in larger challenges. Recall that the panels use community "upvotes" as one signal of concept potential, but also strive to find highly innovative concepts that may have slipped through the

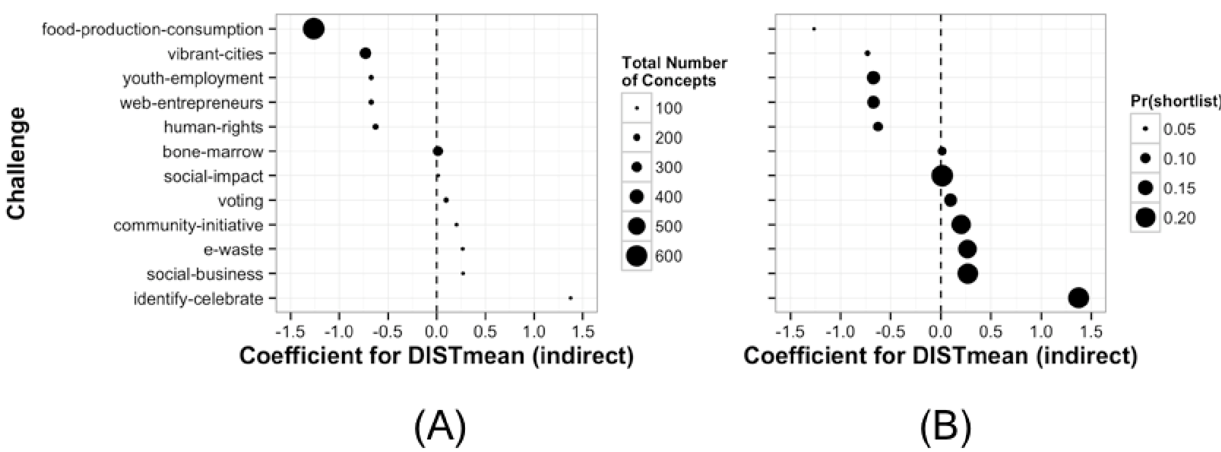

Figure 20. Coefficients for IND-DISTmean estimated by challenge with random slopes models, with 95% CIs, with dots sized by (A) total number of concepts in challenge, and (B) Pr(shortlist) for challenge.

cracks: if concepts built on relatively far indirect sources (controlling for immediate distance) are likely to be more innovative, but potentially less recognizable as such by the community (thereby leading to lower numbers of upvotes), then the expert panel may have an easier time picking out these hidden gems in smaller challenges (where they will not have to sift through as many concepts), and therefore allow for positive effects of distance to emerge. Indeed, challenges with more total concepts (i.e., not just concepts in our analysis sample) tend to have a less positive slope for $IND\text{-}DIST_{MEAN}$, r = –.66 [–.89, –.14], $p < .05$ (see **Figure 20**A). However, it is not clear whether finding hidden gems is that much more effortful (or at least sufficiently more effortful to support this interpretation of the problem variation) for the larger challenges, since (apart from the three large challenges that have N of ~300 or more, i.e., bone-marrow, vibrant-cities, and food-production-consumption), there is relatively little variation among the remaining challenges, with most being around 100-150 concepts.

A related explanation may involve variations in "choosiness": since the expert panel shortlists a relatively fixed number of concepts for each challenge (~20), challenges with more concepts (higher participation) are necessarily more "selective" – that is, the expert panel shortlists a much smaller subset of the submitted concepts. In more selective challenges (with higher levels of participation), the bar for quality might be higher, and the expert panel might be a little more risk averse: in less selective challenges, they may be more willing to risk their selection on ideas that are very novel but perhaps of lesser immediate quality. Indeed, challenges with higher Pr(shortlist) (i.e., less selective challenges) tend to have a more positive slope for $IND\text{-}DIST_{MEAN}$, r = .78 [.37, .94], $p < .01$ (see **Figure 20**B). However, it is not clear how or why lower selectivity might lead to a *positive* (rather than simply neutral) effect of $IND\text{-}DIST_{MEAN}$ We would also need separate novelty and quality ratings of concepts to be able to more directly test the risk aversion explanation.

Leaving aside these statistical or incidental explanations, it could be that there is some other underlying psychologically meaningful challenge-level characteristic that I have not measured that explains the problem variation. For example, Kavadias and Sommer (2009) show by theoretical mathematical analysis that problem *complexity* could moderate the benefits of team knowledge diversity. However, qualitative examination of the challenges did not show obvious variations in problem complexity, particularly in a way that covaried with the effect of $IND\text{-}DIST_{MEAN}$. Nevertheless, it is possible that with a larger sample, and more focused inquiry, complexity might turn out to show a correlation, or other explanatory variables might be discovered.

Overall, I conclude by noting the general concordance of this analysis with the analysis of immediate sources: both analyses find little direct support for the Conceptual Leap Hypothesis, i.e., generally negative estimated effects. From a larger perspective, too, the finding that there might be problem-dependent (or at least problem-varying) positive effects of mean distance of *indirect* sources (bearing in mind the aforementioned statistical caveats) suggests value in incorporating the dimension of temporality (or genealogy, e.g., by distinguishing between immediate and indirect sources) in examining the effects of inspiration sources. Thus, the approach and methodology will be extended to the remaining two chapters, exploring how distance from self and diversity might vary with the immediacy of the sources.

7.0 INDIRECT EFFECTS OF DISTANCE FROM SOLUTION PATH

Having established the possibility of different effects of distance from the problem depending on the “directness” of the sources, I now consider potentially different effects of the distance of indirect sources from one’s solution path. In Chapter 4, I showed that, despite their conceptual distinction, distance from the problem and distance from self showed very similar (i.e., negative) effects on creative success. Nevertheless, the same arguments concerning *precision* of the distance measure might apply when considering distance of *indirect* sources, and it is possible that we might see positive effects of distance of indirect sources from one’s solution path, given the hints at positive effects of problem distance (for some problems) seen in the previous chapter.

7.1 METHODS

The sample (both concepts and indirect sources) for this analysis was the same as in the previous chapter. As with immediate distance from self, we computed 1) $IND\text{-}DISTSELF_{MAX}$, the maximum of the distances of the indirect sources from the root concept, and 2) $IND\text{-}DISTSELF_{MEAN}$, the mean of the distances.

7.2 RESULTS

7.2.1 Descriptive Statistics

Table 18 shows descriptive statistics for the *IND-DISTSELF* predictors, along with their bivariate correlations with the other variables. As before, there are no strong bivariate correlations that hint at possible multicollinearity problems.

7.2.2 Statistical Models

I fit separate sets of models for $IND\text{-}DISTSELF_{MAX}$ and $IND\text{-}DISTSELF_{MEAN}$, with both using the model with controls and $DIST_{MEAN}$ as the baseline model for comparison.

7.2.2.1 Max

Adding $IND\text{-}DISTSELF_{MAX}$ to the best-fitting model (with only controls and $DIST_{MEAN}$) from before results in a small reduction in deviance that also fails a likelihood ratio test of statistical

Table 18: Descriptive statistics for indirect distance from self measures and correlations with other variables

	IND-DISTSELF$_{MAX}$	*IND-DISTSELF*$_{MEAN}$
Descriptives		
Min	–0.60	–0.68
Max	0.00	–0.01
Median	–0.02	–0.15
Mean	–0.05	–0.17
SD	0.08	0.10
Correlations		
SHORTLIST	–0.02	–0.06
FEEDBACK	0.09*	0.04
SOURCESHORT	0.22***	0.11**
DIST$_{MEAN}$	–0.04	0.05
IND-DISTSELF$_{MAX}$		0.63***

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

significance, χ^2 (1) = 1.19, p = 0.27, and a slight increase in the AIC to 413.34 (see **Table 19**). The model estimates a very similar effect to *IND-DIST*$_{MAX}$, i.e., a slightly *negative* effect (estimated with very high uncertainty). Adding a random effect of challenge on *IND-DISTSELF*$_{MAX}$ results in a near-zero estimate of the challenge-level variance in the slope of *IND-DISTSELF*$_{MAX}$, and does not meaningfully decrease deviance from the simpler fixed effects model, χ^2 (2) = 1.06, p = .48 (as before, p-value is halved to correct for overconservativeness),

Table 19: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on *IND-DISTSELF*$_{MAX}$, with comparison to baseline model (controls and *DIST*$_{MEAN}$)

	Baseline model (controls and *DIST*$_{MEAN}$)	*IND-DISTSELF*$_{MAX}$, fixed slope	*IND-DISTSELF*$_{MAX}$ random slope
Fixed effects			
γ_{00}, intercept	–2.78 [–3.43, –2.11]	–2.91 [–3.61, –2.21]	–2.94 [–3.70, –2.17]
γ_{10}, *FEEDBACK*	0.09*** [0.06, 0.12]	0.09*** [0.07, 0.12]	0.09*** [0.07, 0.12]
γ_{20}, *SOURCESHORT*	0.16 [–0.08, 0.39]	0.19 [–0.06 0.43]	0.19 [–0.06 0.44]
γ_{30}, *DIST*$_{MEAN}$	–0.42* [–0.78, –0.06]	–0.43* [–0.79, –0.06]	–0.43* [–0.80, –0.07]
γ_{40}, *IND-DISTSELF*$_{MAX}$		–0.19 [–0.51, 0.13]	–0.20 [–0.55, 0.13]
Random effects			
$u_{0authorj}$	0.13	0.12	0.15
$u_{0challengek}$	0.72	0.69	0.94
$u_{1challengek}$			0.07
Model fit statistics			
Deviance	400.53	399.34	398.29
AIC	412.53	413.34	416.29

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

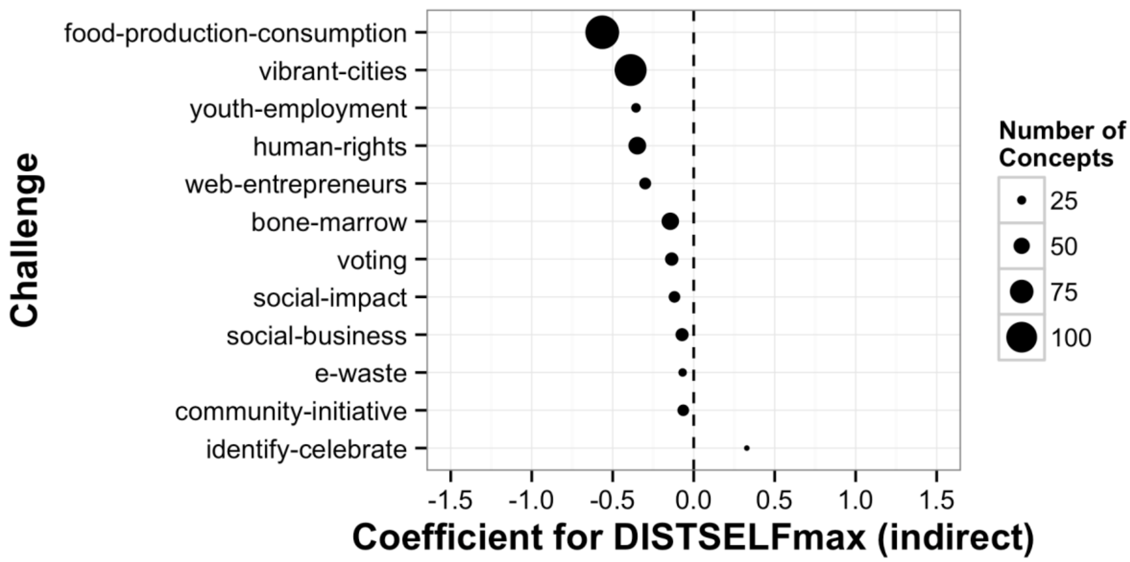

Figure 21. Coefficients for *IND-DISTSELF*$_{MAX}$ estimated by challenge with random slopes models, with 95% CIs, with dots sized by number of sampled concepts in challenge.

and also further increases AIC to 416.29. This, combined with the by-challenge plot of the coefficients (see **Figure 21**), gives us reason to suspect that the mean effect is likely to be either near-zero or else mostly slightly negative (only one challenge seems to be slightly positive).

7.3.2.1. Mean

Similar results are found with *IND-DISTSELF*$_{MEAN}$ as with *IND-DIST*$_{MEAN}$. Adding *IND-DISTSELF*$_{MEAN}$ to the best-fitting model (with only controls and *DIST*$_{MEAN}$) from before results in a small reduction in deviance that also fails a likelihood ratio test of statistical significance, χ^2 (1) = 0.67, p = 0.41, and a slight increase in the AIC to 413.86 (see **Table 20**). However, fitting an additional model with a random effect of challenge on the slope of *IND-DIST*$_{MEAN}$ *did* meaningfully reduce deviance from the fixed slope model, χ^2 (2) = 6.32, p = 0.02 (halved), and estimated a challenge-level variance component of comparable size to the challenge-level

Table 20: Model estimates and fit statistics for cross-classified multilevel logistic regressions of Pr(shortlist) on $\textit{IND-DIST}_{MEAN}$, with comparison to baseline model (controls and $\textit{DIST}_{MEAN}$)

	Baseline model (controls and $\textit{DIST}_{MEAN}$)	$\textit{IND-DIST}_{MEAN}$, fixed slope	$\textit{IND-DIST}_{MEAN}$ random slope
Fixed effects			
γ_{00}, intercept	–2.78 $^{[–3.43, –2.11]}$	–3.02 $^{[–3.92, –2.12]}$	–3.32 $^{[–4.60, –2.03]}$
γ_{10}, *FEEDBACK*	0.09*** $^{[0.06, 0.12]}$	0.09*** $^{[0.06, 0.11]}$	0.09*** $^{[0.06, 0.12]}$
γ_{20}, *SOURCESHORT*	0.16 $^{[–0.08, 0.39]}$	0.17 $^{[–0.07 0.41]}$	0.22 $^{[–0.03 0.47]}$
γ_{30}, $\textit{DIST}_{MEAN}$	–0.42* $^{[–0.78, –0.06]}$	–0.41* $^{[–0.78, –0.05]}$	–0.43* $^{[–0.80, –0.05]}$
γ_{40}, $\textit{IND-DIST}_{MEAN}$		–0.14 $^{[–0.46, 0.18]}$	–0.21 $^{[–0.66, 0.23]}$
Random effects			
$u_{0authorj}$	0.13	0.10	0.22
$u_{0challengek}$	0.72	0.68	3.24
$u_{1challengek}$			0.35
Model fit statistics			
Deviance	400.53	399.86	393.54
AIC	412.53	413.86	411.54

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

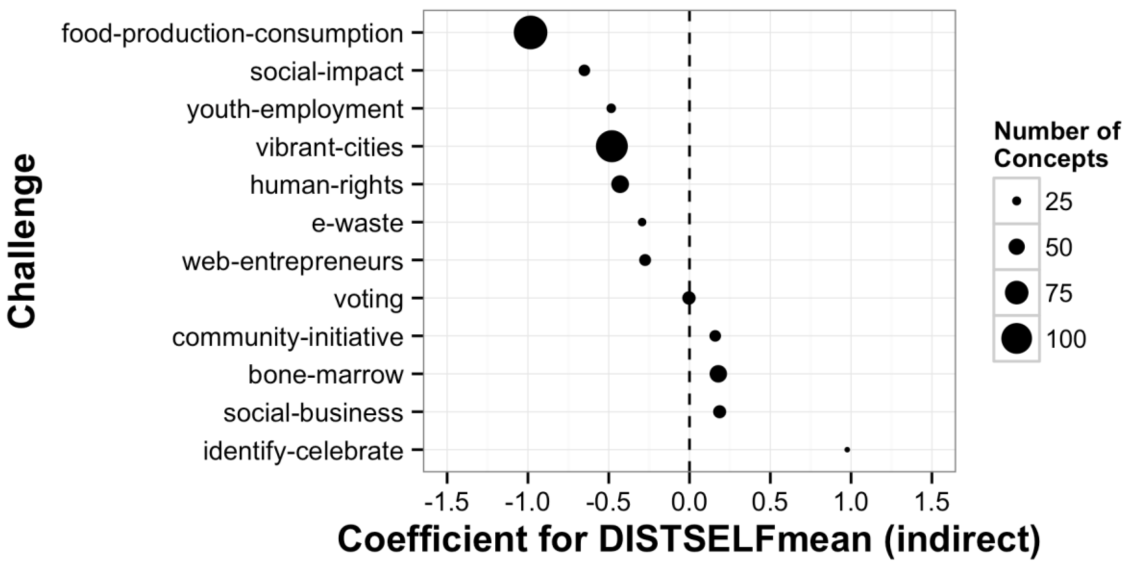

Figure 22. Coefficients for *IND-DISTSELF*$_{\mathrm{MEAN}}$ estimated by challenge with random slopes models, with 95% CIs, with dots sized by number of sampled concepts in challenge.

variance in Pr(shortlist). This random-slopes model's reduction in deviance from the baseline model falls just short of conventional statistical significance, χ^2 (3) = 6.99, p = 0.07, but has a slightly lower AIC (411.54 vs. 412.53). **Figure** 22 shows how, similarly to $IND\text{-}DIST_{MEAN}$, there only seems to be a hint of a positive effect of $IND\text{-}DISTSELF_{MEAN}$ for a subset of the challenges (4 here, compared with 5 from before).

7.3 DISCUSSION

The purpose of this chapter was to explore the possibility that distance of *indirect* sources from one's solution path might show different effects than the distance of immediate sources. The analyses produced substantially similar results to that of the previous chapter: indirect *max* distance from self has a small mean negatively trending slope with no problem variation (with high uncertainty), and there are some hints of problem-variant effects of indirect *mean* distance (although with slightly more precision than before).

However, interpretation of the problem variation is slightly at variance with the previous chapter: as **Figure** 22 showed, the variation seemed to be mostly about the *magnitude* of the negative effect, rather than with the *sign* of the effect. Further, the potential explanations (e.g., statistical precision, judge effort, and judge choosiness) for problem variation with distance from the problem (from the previous chapter) seem to be slightly less applicable here: the correlation of the effect with the number of concepts in the challenge sample was smaller than in the previous chapter, r = –.58 [–.87, –.01], $p < .05$ (compared to r = -.71 from before); correlation with the total number of concepts was r = –.53 [–.85, .06], $p < .10$ (vs. –.66 from before); and the correlation with Pr(shortlist) was r = .53 [–.06, .85], $p < .10$ (vs. .78 from before). Additionally, the fit of the random effects model (with problem variation) seems to be at least slightly better than the fixed effects and controls-only model (with a lower AIC as well). Thus, while the uncertainty in the problem variation effect is still relatively high, we have slightly less reason to believe that it is an artifact of statistical confounds or noise: for indirect sources, then, it seems that increased distance from the solution path may not always negatively impact creative success, and may in some cases even slightly improve it.

It is also worth noting yet again that the negative effect of immediate problem distance remained robust in this analysis, further underscoring its validity. Additionally, the similarity between these results and those of the previous chapter (in terms of finding different patterns of results for immediate vs. indirect sources) further underscores the potential value of examining immediacy of sources as a potential moderator.

8.0 CHAPTER 8: INDIRECT EFFECTS OF DIVERSITY

Having seen potential variations in the effects of source distance by the “directness” of the sources, I now consider the possibility that conceptual diversity may also have different effects when it is “indirect” vs. direct. By similar logic, far combinations may yield novel ideas that require further refinement in order to be useful (and therefore creative): thus, we might expect to see different results when considering the conceptual diversity of *indirect* sources (i.e., sources cited by immediately cited sources), more in line with the Conceptual Leap Combination Hypothesis.

8.1 METHODS

As with indirect distance, I consider indirect sources as sources in levels 2 to 4 of a concept's genealogy. Indirect diversity (here termed *IND-DIV*) was computed the same way as with direct diversity (i.e., by averaging pairwise distances between all indirect inspirations).

The sample for this analysis again was a subset of the initial sample of 707 concepts. To analyze diversity of sources, at least two sources in a set are needed, and not all 707 concepts both cited at least one immediate inspiration *and* at least two indirect inspirations: only 522 concepts met both criteria: the breakdown of this sample by challenge is given in Table 21.

Table 21: Descriptions and number of posts for OpenIDEO challenges in indirect diversity analysis sample

Challenge	Num. concepts *(% diff from previous)*	Num. shortlisted *(% diff from previous)*
Bone-marrow	46 *(35%)*	5 *(29%)*
Community-initiative	32 *(27%)*	7 *(36%)*
E-waste	20 *(23%)*	6 *(25%)*
Food-production-consumption	100 *(32%)*	8 *(20%)*
Human-rights	54 *(13%)*	6 *(14%)*
Identify-celebrate	11 *(54%)*	7 *(46%)*
Social-business	39 *(15%)*	9 *(18%)*
Social-impact	31 *(23%)*	7 *(42%)*
Vibrant-cities	98 *(18%)*	11 *(15%)*
Voting	38 *(19%)*	8 *(0%)*
Web-entrepreneurs	32 *(35%)*	3 *(57%)*
Youth-employment	21 *(34%)*	2 *(33%)*

8.2 RESULTS

8.2.1 Descriptive Statistics

Descriptive statistics and bivariate correlations are given in **Table 22** and **Table 23**. *IND-DIV* does not have any strong correlations with the other predictors, giving little cause for concerns about multicollinearity.

Table 22: Descriptive statistics for indirect diversity measures

Variable	Valid N	Min	Max	Mean	Median	SD
SHORTLIST	522	0	1	0.15	0	0.36
FEEDBACK	522	0	67	9.01	6	10.02
SOURCESHORT	522	0	11	0.67	0	1.06
$DIST_{MEAN}$	522	-2.93	1.67	-0.11	-0.01	0.73
IND-DIV	522	-0.73	-0.02	-0.18	-0.14	0.10

Table 23: Bivariate correlations for indirect diversity measures

Variable	*FEEDBACK*	*SOURCE SHORT*	$DIST_{\mathrm{MEAN}}$	*IND-DIV*
SHORTLIST	0.34***	0.13**	–0.11*	0.04
FEEDBACK		0.11*	–0.01	0.13**
SOURCESHORT			–0.05	0.19***
$DIST_{\mathrm{MEAN}}$				–0.02

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

8.2.2 Statistical Models

Table 24 presents the model estimates and fit statistics for these models. I use a model with the controls variables and $DIST_{MEAN}$ as a baseline for comparing what is added by *IND-DIV*. The baseline model gives a large and statistically significant reduction in deviance compared to the null model, χ^2 (3) = 63.70, p = 0.00. Adding a fixed slope for *DIV* to this model provides a small but marginally significant reduction in deviance, χ^2 (1) = 3.26, p = 0.07, and a slight decrease in the AIC, mitigating concerns about overfitting. The model estimates that a .10 change in *IND-DIV* corresponds to an increase of approximately .45 in the log-odds of being shortlisted. Holding all the other predictors at their mean values, changing from an *IND-DIV* of –0.20 (close to the mean value in the sample) to –0.10) increases Pr(shortlist) from 0.13 to 0.19. **Figure 22** plots this estimated effect of *IND-DIV* in relation to observed Pr(shortlist) for 4 equal N bins

Table 24: 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 and *DIST*$_{MEAN}$)	With *IND-DIV*, fixed slope	With *IND-DIV* random slope
Fixed effects			
γ_{00}, intercept	–2.80 $^{[-3.44,\ -2.16]}$	–1.98 $^{[-3.10,\ -0.86]}$	–2.12 $^{[-3.10,\ -0.86]}$
γ_{10}, *FEEDBACK*	0.09*** $^{[0.06,\ 0.12]}$	0.09*** $^{[0.07,\ 0.12]}$	0.09*** $^{[0.07,\ 0.12]}$
γ_{20}, *SOURCESHORT*	0.16 $^{[-0.08,\ 0.39]}$	0.12 $^{[-0.12,\ 0.35]}$	0.12 $^{[-0.12,\ 0.35]}$
γ_{30}, *DIST*$_{MEAN}$	–0.44* $^{[-0.82,\ -0.07]}$	–0.45* $^{[-0.83,\ -0.06]}$	–0.45* $^{[-0.83,\ -0.06]}$
γ_{40}, *IND-DIV*		0.45^{m} $^{[-0.04,\ 0.94]}$	0.34^{m} $^{[-0.04,\ 0.94]}$
Random effects			
$u_{0authorj}$	0.12	0.13	0.12
$u_{0challengek}$	0.60	0.88	1.35
$u_{1challengek}$			0.03
Model fit statistics			
Deviance	372.65	369.39	369.13
AIC	384.65	383.39	387.13

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

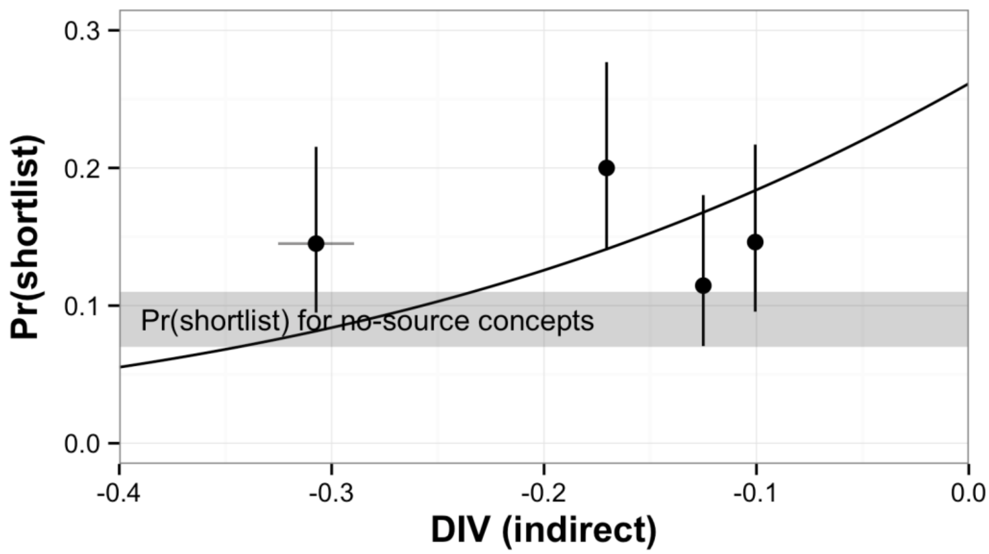

Figure 22. Model-fitted and observed relationship between *IND-DIV* and Pr(shortlist). Fitted values evaluated at mean values of feedback, source shortlist, and $DIST_{MEAN}$. Observed points are in equal N quartile bins. Vertical and horizontal error bars for points are 95% CI for Pr(shortlist) and *IND-DIV*.

(with 95% CIs). Although the observed points suggest a potential quadratic relationship, note the relatively wide CIs for Pr(shortlist), and also note that plotting *IND-DIV* against the residuals from the fixed slope model does not suggest that a quadratic term would add anything. Indeed, a model fitted with a quadratic term for *IND-DIV* does no better in a likelihood-ratio test vs. the baseline model, χ^2 (2) = 3.52, p = 0.17, and also results in a higher AIC than the model with just a linear term for *IND-DIV* (385.12 vs. 383.39), suggesting that any additional deviance reduction may be due to overfitting.

The estimated positive effect of *IND-DIV* did not appear to vary by challenge. Fitting an additional model with a random slope for *IND-DIV* estimates essentially zero challenge-variance, does not meaningfully decrease variance from the fixed slope model, χ^2 (2) = 0.26, p = .44 (p-value is halved, as before), and also further increases AIC.

8.3 DISCUSSION

In this chapter, I extended the approach and methodology from the previous two chapters, considering how the effects of diversity of sources might be different depending on the sources' immediacy. Analyzing diversity of indirect sources indeed yields different results than direct diversity; specifically, the extremely small positive effect of direct diversity is now significantly amplified, to a comparable magnitude as the effect of $DIST_{\mathrm{MEAN}}$ (albeit with considerable uncertainty about the true size of the effect). Thus, the Conceptual Combination Hypothesis appears to find at least partial support when considering indirect rather than direct sources: far combinations do appear to benefit one's creative success, but only if they are "indirect" (i.e., sources of one's sources).

One potential concern with this analysis is that I did not control for the number of cited inspirations: intuitively, the more inspirations are cited, the higher the likelihood for diverse pairs to exist in the set. Perhaps it is the number of indirect inspirations in a set and not the *diversity* of those inspirations that matters: it could be that number of indirect inspirations (which can be thought of as the size of one's ego network in a given window) indicates being in a "popular" region of the design space, which in turn is correlated with creative success. But this concern is not borne out in the data. While indirect diversity and the number of indirect inspirations are indeed correlated ($r = .33$ [0.27, 0.39], $p < .001$), including the number of indirect inspirations in the final model does not substantively change the estimate of the effect of indirect diversity ($B = 0.46$ [–0.02, 0.94]).

Again, it is worth noting the continued robustness of the negative effect of problem distance for immediate sources. Also, while moving from immediate to indirect sources increased the magnitude of the estimated effect of source diversity (in contrast to revealing potential problem variation for the effects of distance), these results are similar to that of the preceding chapters in that all three chapters demonstrate the potential importance of considering the immediacy of inspiration sources as a potential moderator of their effects.

9.0 GENERAL DISCUSSION

I now conclude this dissertation by summarizing the insights gained and working out their implications for the theory and practice of creative inspiration.

9.1 SUMMARY OF FINDINGS

This dissertation was conducted to aid in the discovery of principles that can guide the curation and use of sources of inspiration in the creative process. I found two broad related but distinct categories of recommendations commonly found in the literature, which formed the basis for the two main research questions addressed in this dissertation:

3) **What are the relative benefits of different levels of source conceptual distance for creative outcomes?**
4) **What are the relative benefits of different levels of source conceptual combination distance for creative outcomes?**

Question 1 led to testing of the Conceptual Leap Hypothesis, which postulates that the best creative insights come from high levels of source conceptual distance. I designed this study to address some key potential reasons for mixed empirical findings in prior work, namely time scale, statistical power, and problem variation. I also examined two operationalizations of distance: distance from the problem, and distance from self (i.e., the solution path). Question 2 led to testing of the Conceptual Combination Hypothesis from prior literature, which posits that better creative insights accrue from higher levels of conceptual combination distance (or higher diversity of sources). This hypothesis was relatively well supported for novelty of results, but not as well for the combination of novelty and quality. Cutting across these two questions, I also wondered about how different levels of immediacy of sources might moderate the effects of these variables: can we trust far sources or combinations to generate immediately usable breakthrough ideas, or do they have to be combined with some other process to produce good ideas (e.g., refinement/iteration strategies)?

Figure 23 summarizes the main findings from this dissertation. For the first major question, this dissertation yielded strong evidence *against* the Conceptual Leap Hypothesis: far from being a consistent benefit, conceptual distance was found to often be harmful to creative success. The most robust finding was that preferring to directly cite sources that are conceptually far from one's problem domain resulted in consistently *worse* creative success: the effect, while estimated with some imprecision, was robust across the different problems on the platform (see **Figure 24** for a summary of problem variation across the different variables), robust across different parameter settings for the computational distance measures, and remained unchanged with the addition of distance from self and diversity measures. This negative effect was muted when considering the distance of indirect sources, with consistently negative but much smaller overall effects (i.e., averaged across problems) for both distance from the problem and distance from self. There were some hints of positive effects for some problems when considering only the *mean* distance from the problem; however, interpretation of these findings is uncertain given

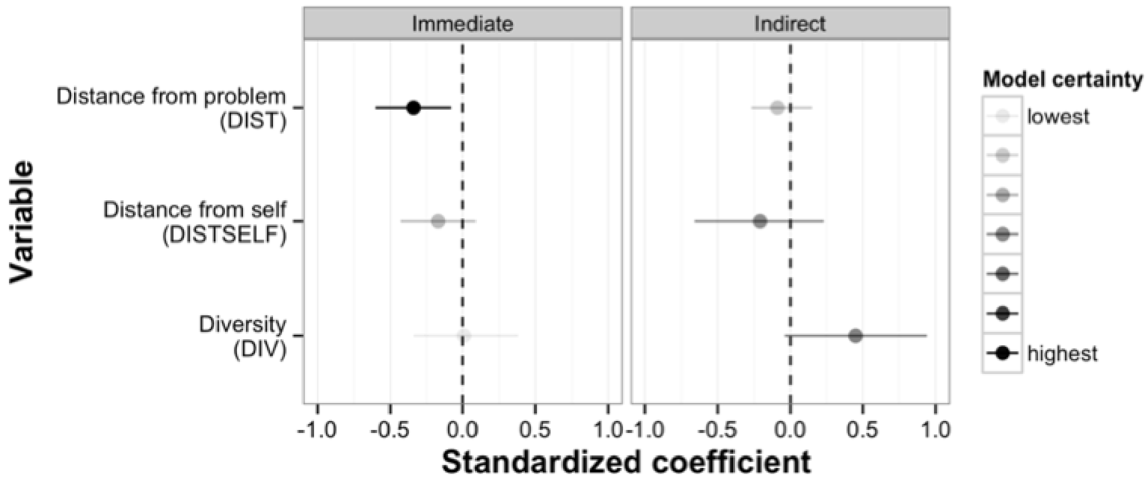

Figure 23. Summary of effects by variable and source type. Coefficients are standardized by multiplying them by the variable's standard deviation. Model certainty is approximated by calculating how much lower the Akaike Information Criterion (AIC) of the best fitted model with the variable in question and the AIC of its baseline model (controls only for immediate problem distance, controls plus immediate distance for all other models). For reference, the best-fitting model for DIST had an AIC that was lower than its baseline by 4.26 points; in contrast, AIC for the best-fitting model for DIV was *higher* than its baseline by 2 points.

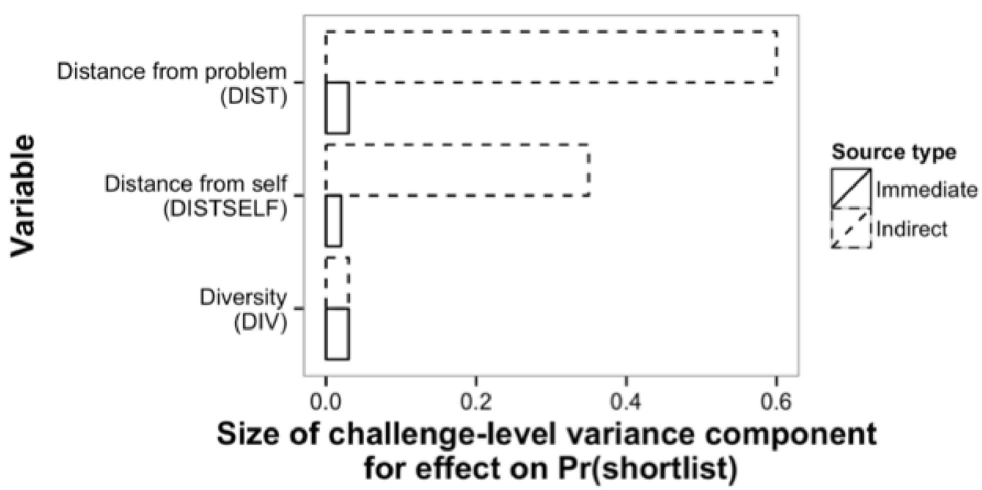

Figure 24. Summary of degree of problem variation by variable and source type.

the high correlations with potential nuisance variation (e.g., sample size), and lack of available data to test possible explanations for the problem variation.

With respect to the second major question, partial support was found for the Conceptual Combination Hypothesis: controlling for mean problem distance of direct sources (along with other control variables), there was a very small (but statistically insignificant) estimated positive effect of conceptual combination distance of *direct* sources; this positive effect grew larger when considering *indirect* sources, to a similar size as the negative effect of mean problem distance, but with more uncertainty, and only a marginally significant improvement in model fit when adding a fixed effect of indirect diversity.

9.2 CAVEATS AND LIMITATIONS

Some caveats and limitations should be discussed before addressing the implications of this dissertation. First, the statistical patterns observed here are conditional: i.e., I find that increased conceptual distance of *cited* inspiration sources negatively impacts Pr(shortlist). My data is silent on the effects of distance for concepts that did not cite sources. However, these concepts were overall of lower quality; thus, it is unlikely that the negative effects of distance are due to attrition (e.g., beneficial far inspirations not being observed). Nevertheless, we should 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, the Conceptual Leap and Conceptual Combination hypotheses may be more applicable to conscious inspiration processes (e.g., analogy, for which conscious processing is arguably an important defining feature; Christensen & Schunn, 2005).

Relatedly, some might be concerned about the meaningfulness of citations on this platform: my inference from this data that the use of different kinds of sources can impact creative success depends on the assumption that the sources were actually being used in the development of the concept. This assumption may not actually be warranted: for instance, citations may mainly reflect attempts to give an appearance of quality or primarily serve a social function (e.g., gaining attention to solicit feedback).

There are at least two sources of evidence that address this potential concern. First, in many concepts, the authors do mention not only *that* a source inspired their idea, but also *how*. For example, in “E-Life Cycled” (Reader, 2012), a concept for the e-waste challenge focusing on creating a brand/label for new products manufactured substantially with reprocessed materials (as a way to set a business’s products apart from the competition), the author states that he was “[b]uilding on [name redacted]’s original inspiration for a Brand for "Made from E-Waste" and incorporating [name redacted]’s "E-Life" notion (raised initially in [name redacted]’s Pure Tech concept) for renewing and reinvigorating materials from end-of-original-life products…I simply felt the concept [name redacted] envisioned was too valuable to leave out of the contributions in this concepting phase.” The original inspiration “Made from E-waste” proposed a branding campaign modeled after the success of an electronic chip-making company’s (Intel’s) successful “Intel Inside” branding campaign. As another example, in “Farmers' Market (e)waste not Booth” (Shu, 2012), a concept about setting up education and recycling booths at farmer’s markets, the author builds on an insight about human behavior from an inspiration about battery collection efforts in Germany, stating, “[g]iven the popularity of farmers’ markets, they’d be great places to set up an interactive and informative booth where people could learn about electronics recycling and reusing, as well as reducing their electronics consumption. As [name redacted] pointed out in her Battery Collection inspiration post, it's easier for people to drop off e-waste at a location that they already regularly visit - therefore the booth would also serve as a drop-off spot for unwanted electronics.” These two examples illustrate how citations to inspirations are meaningful indicators of substantive shaping of concepts by the inspirations they cite. Secondly, as we have seen, concepts that cite inspirations have higher creative success than concepts that do not cite inspirations, controlling for feedback, providing additional evidence that citations are meaningfully involved in the development of concepts, and that observed effects of citations are not due to them serving a “social function” (e.g., spurring more attention and feedback, which is what actually improves the concept).

A second potential caveat is that I have not directly measured novelty here. Conceivably, the benefits of distance or diversity may only be best observed for the novelty of ideas, and not necessarily quality, consistent with some recent work (Franke, Poetz, & Schreier, 2013). However, novelty *per se* is not innovation; I contend that to fully understand the effects of distance on design innovation, we must consider its impacts on both novelty and quality together (as our shortlist measure does). Further, concerns about risk aversion (preferring feasibility over novelty) as an explanation for the negative effects of distance are mitigated when considering that there were estimated positive effects of source diversity, which theoretically also accrues benefits via increasing novelty (e.g., through generation of emergent features from the combinations).

Related to this, there is a potential concern over the binary nature of the creative outcome measure: perhaps both near- and far-inspired ideas get past the binary threshold for being shortlisted, but the absolute quality or (long-term, eventual) impact of far-inspired ideas will far exceed that of near-inspired ones. Under this logic, the Conceptual Leap Hypothesis may still be right, if far-inspired ideas produce both worse and better quality ideas, thus sending fewer ideas into the shortlist stage, but the ultimate impact of those that do make it past the threshold ends up dwarfing the impact of any of the other more "mundane" near-inspired ideas. The present data cannot address this caveat: finer-grained expert ratings of the novelty and quality of ideas are needed to address this potential alternative explanation for this dissertation's findings. These finer-grained ratings will allow us to determine if far-inspired ideas are, on average, of much higher quality or novelty than near-inspired ideas, regardless of the fact that more near-inspired ideas make it past the shortlist threshold.

9.3 IMPLICATIONS AND FUTURE DIRECTIONS

9.3.1 Alternate Pathways to Good Ideas

These caveats notwithstanding, this dissertation yields some useful insights for the literature. First, my results do stand in opposition to the Conceptual Leap Hypothesis. In tandem with prior opposing findings (reviewed in the introduction), my 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 innovation.

I should be clear that my findings do not imply that *no* innovative ideas come from far sources; rather, our data suggest that *overreliance* on far sources (e.g., as indicated by a high mean level of distance) negatively impacts ideation (perhaps due to cognitive costs that might not be mitigated by extra processing time; Perkins 1997). However, my findings do suggest that highly innovative ideas can often come from relying almost not at all on far sources. These good ideas may arise from iterative, deep search, a mechanism for innovative 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 innovation.

It is worth noting that there are potential discrepancies between the way distance is conceptualized and measured in this dissertation, and the notion of "far sources" in the analogy literature, which focus on the joint property of having some base level of *structural* similarity and low *surface* similarity. To be most precise, the claims in the theoretical analogy literature about the benefits of far sources (e.g., in Gentner & Markman, 1997; Ward, 1998) are about this particular sort of far sources. However, while I do not explicitly measure structural and surface similarity separately, I argue that my data are not completely disconnected from these theoretical claims, and can in fact inform the assessment of these claims. First, these theories have not clearly specified how to distinguish between structural and surface similarity in complex domains such as design and social innovation: what is "surface" in one mapping (e.g., the shape of a logo when mapping the insight of using logos to evoke branding) may be "structural" in another mapping (e.g., transferring the use of a particular shape pattern to evoke a particular message). In the absence of clear indications of the nature of each mapping, and clear principles for distinguishing structural from surface similarity across a diversity of source-target mappings, it seems reasonable to use "overall similarity" (as tracked by LDA) to measure distance, to allow for the diversity of possible structural mappings. Second, overall similarity of source and target can be a clue to the potential for structural alignment; indeed, Gentner (2010) has argued that children use such cues to "bootstrap" their development of relational and analogical reasoning ability. Finally, arguably in most cases in this dataset where inspirations were cited, some mapping was found to the problem, and so there is at least a base level of structural alignment present (i.e., it would be difficult to argue that the majority of inspiration citations reflected *only* superficial feature transfer); given this, it seems reasonable to say that comparisons between inspirations in terms of distance from the problem would largely track surface similarity, thereby aligning the LDA measure of distance more closely to the conceptualization of analogical distance (i.e., has structural similarity AND low surface similarity) than one might initially suppose. For these reasons, I argue that the findings in this dissertation about the negative effects of distance (and conversely the positive effects of conceptually near sources) are relevant for revisiting theories of analogical distance and its role in creativity.

9.3.2 Moderators and Enablers of Conceptual Distance Effects

9.3.2.1 Conceptual Distance of Ideas vs. People

Rather than overturning the Conceptual Leap Hypothesis in light of the present data, it may be fruitful to consider how it might be revised/supplemented with specifications of enabling conditions and contextual moderators of the benefits of conceptual distance. One potential enabling condition is suggested by reflections on tensions between the current work and research that has shown the importance of interdisciplinarity for breakthrough innovation. For example, a number of studies have shown boosts in innovation from collaborations between problem solvers from different disciplines and diverse expertises (Bercovitz & Feldman, 2011; Ruef, 2002; Singh & Fleming, 2009; Taylor & Greve, 2006; Uzzi & Spiro, 2005), and some other recent studies have shown that problem solvers from outside the problem domain can often produce the most creative solutions to the problem (Franke et al., 2013; Jeppesen & Lakhani, 2010). Perhaps there is a critical distinction between conceptual distance of *ideas* vs. conceptual distance of *people*. Returning to our reflections on the potential costs of processing far sources, we suggest that expertise in the distant source domain may be a crucial mediator of its benefits. In interdisciplinary collaborations, the expertise of each actor might bypass the cognitive costs of deeply understanding the far domain, and filter 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. Such studies would be highly informative for innovation theory, and also have potential practical implications: if ideas from other domains do in fact have a unique connection to creative breakthroughs, but only when they come from experts in those domains, then resources for finding cross-domain sources may be better routed to finding cross-domain *collaborators*.

9.3.2.2 Problem Variation

Another potential moderator is problem characteristics. This study provided partial evidence that there might be problem variation for the distance of indirect sources: however, this variation was correlated with a potential confound of selectivity and/or statistical power. Nevertheless, in light of prior work showing problem variation of stimuli effects in design ideation (Chiu and Shu, 2012; Goldschmidt & Smolkov, 2006; Liikkanen & Perttula, 2008), it may be fruitful to further examine problem variation. As mentioned earlier, one potentially important dimension of variation is problem complexity. It could be that as problem domains increase in complexity, specialization might also increase, as the "burden of knowledge" becomes too great for any one person or team to carry (Jones, 2009). In this situation, good ideas might become "trapped" or "siloed" in different disciplines, making it more important for innovators to draw from outside their discipline in order to create good ideas. Partial support for this conjecture comes from the literature on interdisciplinary team innovation, and from social network theories of innovation that 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. Again, however, these theories and findings might only apply to distant *people*, and not *ideas per se*, given potential cognitive costs.

9.3.2.3 Source Processing Strategies

Finally, it would be interesting to examine potential moderating influences of source processing *strategies*. In my data, closer sources were more beneficial, but good ideas also did come from far sources; however, as I 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 near 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); can this strategy also be extended to leverage far sources? Are there other processing strategies that expert creative designers apply uniquely to far sources (e.g., to deal with potentially un-alignable differences)? Answering these questions can shed further light on the variety of ways designers can be inspired by sources to produce innovative design ideas.

9.3.3 Immediate vs. Indirect Effects of Inspiration Sources

This dissertation also demonstrated the potential value of distinguishing between immediate and indirect sources. Perhaps owing to the nature of the creative outcome measure (which combines considerations of both novelty and quality), this dissertation suggested that far conceptual combinations may, on average, directly generate ideas that are *slightly* better than average (although our statistical confidence in this is very low), but that these ideas may then go on to fuel even better ideas. This pattern of results suggests that far conceptual combinations may be good for generating novel, but not necessarily *immediately* feasible/useful ideas – these ideas may need further processing or refinement before they can be considered “good ideas” and potentially make a meaningful contribution to the problem at hand.

It is also worth noting that different results were found for immediate vs. indirect source distance, with indirect source distance appearing to potentially be helpful for some problems, whereas it was consistently harmful when considering immediate sources. These results suggest that further investigations should, if possible, consider the effects of different source characteristics or processing strategies (or, more generally, ideation strategies/methods) across the *phases* of the creative process: for example, far combinations may be especially helpful for expanding the idea space (divergent processes), but less helpful for iterative, deep search, or finding improvements for existing ideas (convergent processes).

9.4 SUMMARY OF CONTRIBUTIONS

I conclude by reviewing the contributions of this dissertation to the literature. In this dissertation I have:

1. Developed and validated a computational methodology for studying conceptual distance with complex design concepts. This methodology addresses significant issues of efficiency and scalability faced in prior work: some of the analyses conducted in this dissertation (e.g., distance from self, pairwise distances between sources) would have been costly to the point of intractability without the methodology.
2. Challenged the widespread but unevenly supported notion that far sources provide the best insights for creative ideation; instead, I have shown that overreliance on far sources can harm ideation success, and that good ideas can often come from very near sources. Combined with the weight of prior similar findings of neutral or negative effects of distance, and the specific design features of the current study (namely addressing issues of time scale, statistical power, and problem variation), this dissertation helps the

literature converge on a more confident conclusion that the Conceptual Leap Hypothesis may need to be overturned, or at least revised/refined.

3. Discovered the potential value of incorporating a temporal dimension into analyses of the impact of sources of inspiration. I find evidence of differential impacts of source distance and diversity (viz., increased problem variation for the effect of source distance, and a more robust positive effect of source diversity) when considering sources farther back in ideas' conceptual genealogies.

It is my hope that these contributions will inspire further research that continues to enrich and deepen the cognitive science of creativity and innovation.

APPENDIX: TOPIC MODELING TECHNICAL DETAILS

This appendix presents technical details for my topic model-building approach. Recall that 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, Gerrish, Wang, Boyd-graber, & Blei, 2009).

I 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 25** shows the mean fit (with both continuous and binary similarity judgments) at each level of K.

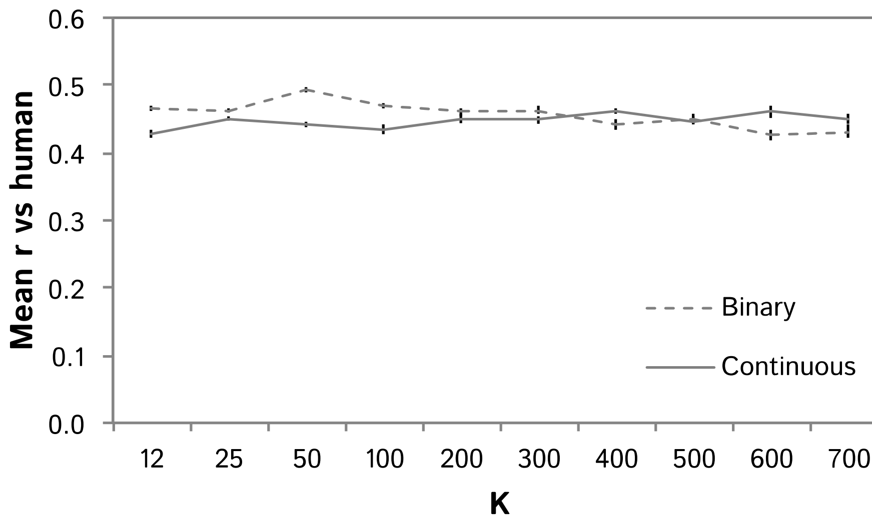

Figure 25. Mean fit (with ±1 SE) vs human judgments for LDA cosines by level of K.

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 I wanted to optimize for fit with human judgments of conceptual distance overall, I 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, I 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 I reported in section 3.2.2.3, the main results of this dissertation show robustness to different settings of K.

BIBLIOGRAPHY

Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of Experimental Psychology*, *44*(4), 288-291.

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

Baruah, J., & Paulus, P. B. (2011). Category assignment and relatedness in the group ideation process. *Journal of Experimental Social Psychology*, *47*(6), 1070-1077.

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

Baughman, W. A., & Mumford, M. D. (1995). Process-analytic models of creative capacities: Operations influencing the combination-and-reorganization process. *Creativity Research Journal*, *8*(1), 37-62.

Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, *40*(1), 81-93.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious einstellung (set) effect. *Cognition*, *108*(3), 652-661.

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., & Lafferty, J. (2006). Correlated topic models. In *Neural information processing systems.*

Blei, D., & Lafferty, J. (2007). A correlated topic model of Science. *The Annals of Applied Statistics, 1*(1), 17-35.

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

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

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376. doi: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.

Chan, J., & Schunn, C. (2014). The impact of analogies on creative concept generation: Lessons from an in vivo study in engineering design. *Cognitive Science.*

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).

Chase, C. C., Herman, A., & Dow, S. P. (2012). *Exploring variation in examples for design instruction: Similarity fosters learning, diversity promotes creativity* [Technical Report].

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

Christensen, B. T. & Schunn, C. D. (2005). Spontaneous access and analogical incubation effects. *Creativity Research Journal*, *17* (2-3), 207-220.

Chrysikou, E. G., & Weisberg, R. W. (2005). Following the wrong footsteps: Fixation effects of pictorial examples in a design problem-solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1134-48.

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..

Dyer, J. H., Gregersen, H. B., & Christensen, C. M. (2011). *The innovator's DNA: Mastering the five skills of disruptive innovators.* Boston, MA.

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.

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.

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.

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.

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.

Gielnik, M. M., Frese, M., Graf, J. M., & Kampschulte, A. (2011). Creativity in the opportunity identification process and the moderating effect of diversity of information. *Journal of Business Venturing*, *27*(5), 559-576.

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*, (ahead-of-print), 1-18.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752-775.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences*, *6*(1), 3-40.

German, T. P., & Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. *Psychological Science*, *16*(1), 1-5.
Goldschmidt, G., & Smolkov, M. (2006). Variances in the impact of visual stimuli on design problem solving performance. *Design Studies*, *27*(5), 549-569.
Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proc Natl Acad Sci U S A*, *101 Suppl 1*, 5228-35. doi:10.1073/pnas.0307752101

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, *15*(1), 55-71.

Hampton, J. A. (1997). Emergent attributes in combined concepts. In T. B. Ward, S. M. Smith, & J. Viad (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 83-110). Washington, DC: American Psychological Association.

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

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly, 42*(4), 716.

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.

Henry, T. (2011). *The accidental creative : How to be brilliant at a moment's notice.* New York: Portfolio/Penguin.

Herring, S. R., Chang, C. C., Krantzler, J., & Bailey, B. P. (2009). Getting inspired! Understanding how and why examples are used in creative design practice. In *Proceedings of the 27th international conference on human factors in computing systems* (pp. 87-96).

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

Howard-Jones, P. A., Blakemore, S. -J., Samuel, E. A., Summers, I. R., & Claxton, G. (2005). Semantic divergence and creative story generation: An fMRI investigation. *Cognitive Brain Research, 25*(1), 240-250.

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. In *Computational information retrieval* (pp. 121-144). SIAM, Philadelphia.

Johansson, F. (2006). *The Medici effect: What elephants & epidemics can teach us about innovation.* Boston, MA: Harvard Business Press.

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. doi:10.1111/j.1467-937X.2008.00531.x

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.

Kavadias, S., & Sommer, . C. (2009). The effects of problem structure and team diversity on brainstorming effectiveness. *Management Science*, *55*(12), 1899-1913.

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

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*(6), i-95.

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

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, *12*(2), 181-194.

Marsh, R. L., Bink, M. L., & Hicks, J. L. (1999). Conceptual priming in a generative problem-solving task. *Memory & Cognition*, *27*(2), 355-363.

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1996). How examples may (and may not) constrain creativity. *Memory & Cognition*, *24*(5), 669-680.

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*, n/a. doi:10.1002/jocb.49

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

Mobley, M. I., Doares, L. M., & Mumford, M. D. (1992). Process analytic models of creative capacities: Evidence for the combination and reorganization process. *Creativity Research Journal*, *5*(2), 125-155.

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.

National Academy of Engineering (2005). *Educating the engineer of 2020: Adapting engineering education to the new century.* Washington, D.C.: The National Academies Press.

Nijstad, B. A., Stroebe, W., & Lodewijkx, H. F. M. (2002). Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of Experimental Social Psychology*, *38*(6), 535-544.

Öllinger, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem solving. *Experimental Psychology*, *55*(4), 269-282.

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.

Perttula, M., & Sipilä, P. (2007). The idea exposure paradigm in design idea generation. *Journal of Engineering Design*, *18*(1), 93-102.

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

Purcell, A. T., & Gero, J. S. (1992). Effects of examples on the results of a design activity. *Knowledge-Based Systems*, *5*(1), 82-91.

Purcell, A. T., & Gero, J. S. (1996). Design and other types of fixation or is fixation always incompatible with innovation? *Design Studies*, *17*(4), 363-383.

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/

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

Reader, P. (2012, August 14). *E-Life Cycled*. Retrieved from http://openideo.com/challenge/e-waste/concepting/e-life-cycled

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.

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

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

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

Shu, C. (2012, July 24). *Farmers' market "(e)waste not" booth.* Retrieved from http://openideo.com/challenge/e-waste/concepting/farmers-market-ewaste-not-booth

Simonton, D. K. (1988). *Scientific genius: A psychology of science.* New York, NY: Cambridge University Press.

Singh, J., & Fleming, L. (2009). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, *56*(1), 41-56.

Smith, S. M., Kohn, N. W., & Shah, J. (2008). What you see is what you get: Effects of provocative stimuli in creative invention. In *Proceedings of NSF international workshop on studying design creativity*.

Smith, S. M., Ward, T. B., & Schumacher, J. S. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*(6), 837-45.

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.

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.

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* (3 ed.). New York, NY.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, *111*(2), 447-504.

Vedres, B., & Stark, D. (2010). Structural folds: Generative disruption in overlapping groups. *American Journal of Sociology*, *115*(4), 1150-1190.

Vogel, C. M., Cagan, J., & Boatwright, P. B. H. (2005). *The design of things to come: How ordinary people create extraordinary products.* Upper Saddle River, NJ.

von Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy, 34*(10), 1591 - 1607.

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.

Ward, T. B. (2008). The role of domain knowledge in creative generation. *Learning and Individual Differences*, *18*(4), 363-366.

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. doi: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.

Wilkenfeld, M. J. (1995). *Conceptual combinations: Does similarity predict emergence?* Unpublished master's thesis, Texas A&M University, College Station, TX.

Wilkenfeld, M. J., & Ward, T. B. (2001). Similarity and emergence in conceptual combination. *Journal of Memory and Language*, *45*(1), 21-38.

Wisniewski, E. J. (1997). Conceptual combination: Possibilities and esthetics. In T. B. Ward, S. M. Smith, & J. Viad (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 51-81). Washington, DC: American Psychological Association.

Wisniewski, E. J., & Gentner, D. (1991). On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. In G. B. Simpson (Ed.), *Advances in Psychology: Understanding word and sentence* (Vol. 77, pp. 241 - 284). North-Holland.

Youmans, R. J. (2011). The effects of physical prototyping and group work on the reduction of design fixation. *Design Studies*, *32*(2), 115-138.

Yu, L., Kraut, B., and Kittur, A. (2014). Distributed analogical idea generation: innovating with crowds. *In Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'14)*.

Zahner, D., Nickerson, J. V., Tversky, B., Corter, J. E., Ma, J., Jeffrey, J., . . . JingMa. (2010). A fix for fixation? Rerepresenting and abstracting as creative processes in the design of information systems. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *24*(02), 231-244.

Zeger, S. L., Liang, K. -Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 1049-1060.

Zeng, L., Proctor, R. W., & Salvendy, G. (2011). Fostering creativity in product and service development: Validation in the domain of information technology. *Human Factors, 53*(3), 245-70.