
Conceptual Distance Matters When Building on Others' Ideas in Crowd-Collaborative Innovation Platforms

Joel Chan

PhD Candidate, Learning
Research and Development
Center
University of Pittsburgh
3939 O'Hara St
Pittsburgh, PA 15260, USA
joc59@pitt.edu

Steven P. Dow

Assistant Professor, Human
Computer Interaction Institute
Carnegie-Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213, USA
spdow@cs.cmu.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
CSCW'14 Companion, Feb 15-19 2014, Baltimore, MD, USA
ACM 978-1-4503-2541-7/14/02.
<http://dx.doi.org/10.1145/2556420.2556500>

Christian Schunn

Professor, Learning Research
and Development Center
University of Pittsburgh
3939 O'Hara St
Pittsburgh, PA 15260, USA
schunn@pitt.edu

Abstract

In crowd-collaborative innovation platforms, other contributors' ideas can serve as sources of inspiration for creative ideas, but what patterns of interactions with others' ideas are most helpful? We investigate the hypothesis that building on inspiration sources that are conceptually far from one's target domain are most helpful, a popular hypothesis with mixed empirical support. We predict the success rate of 2,344 ideas for 12 different design challenges in a collaborative Web-based innovation platform based on their cited sources' conceptual distance from the target domain (measured using probabilistic topic modeling of the ideas). Surprisingly, we find that innovators who cite conceptually near sources of inspiration achieve a higher success rate than those who prefer far sources. We discuss implications for research and development of crowd-collaborative innovation platforms.

Author Keywords

Creativity; crowdsourcing; collaboration; design; analogy

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

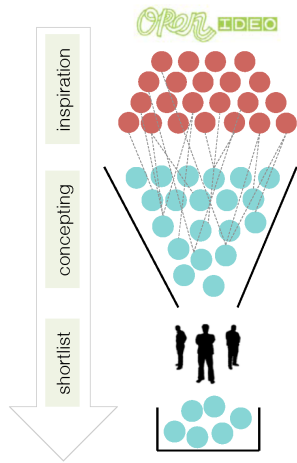


Figure 1. Illustrated OpenIDEO process: contributors post inspirations (top), which are cited in concepts (middle), a subset of which are shortlisted by an expert panel for further development.

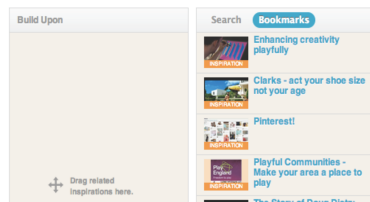


Figure 2. Screenshot of citation interface: while posting/refining concepts, contributors drag and drop inspirations (bookmarked or searched for) to the left column to cite them as sources.

Introduction

Finding and building on sources of inspiration is part of any creative process [1] and often an important contributor to creative breakthroughs [2]. In crowd-collaborative innovation platforms (e.g., Quirky.com, OpenIDEO.com) where many people contribute ideas and collaborate with each other to solve a variety of creative problems, there is an opportunity to shape these platforms so that previously contributed ideas serve as sources of inspiration for further ideation. To better understand how innovators interact with other ideas, we need more empirical data on what sorts of ideas inspire novel and high-quality concepts. One hypothesis from the creativity literature is that candidate sources with the highest potential for inspiring creative breakthroughs are those that are conceptually far from one's working domain [3], i.e., structurally similar ideas with many surface (or object) dissimilarities (e.g., atom/solar system analogy).

The empirical evidence for this hypothesis is mixed: a number of studies have shown an advantage of far over near sources for creative outcomes [4-5], while others sources for creative outcomes [6], or even an advantage of near over far [7]. These inconsistent results could stem from the tendency to observe only short time-slices of the creative process (e.g., ~30-60 mins), whereas more time and iteration (e.g., over the course of days/weeks) may be necessary to benefit from far sources, due to the cognitive challenges of mapping far sources [4]. Statistical power has also been an issue, with most studies having an N of 12 or less per treatment cell, insufficient to detect even medium to large-sized statistical effects.

Methods

Overview

We investigate this interesting but unevenly supported hypothesis, and address prior methodological limitations by studying large numbers of ideas (on the order of thousands) and at a realistic time scale (days/weeks), in the context of OpenIDEO (www.openideo.com), a large-scale Web-based crowd-sourced innovation platform that addresses social innovation problems (e.g., managing e-waste, increasing accessibility in elections). Over ~10 weeks, contributors to the platform first 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, and then *concepts*, i.e., specific solutions to the problem, a subset of which are shortlisted for further development (see Fig. 1). Concepts are typically ~150 words long, providing more detail than one or two words/sentences/sketches, but less detail than a full-fledged design report (see Fig. 4). Contributors are encouraged to build on others' ideas: when posting concepts, contributors are prompted to cite inspirations that serve as sources of inspiration for their idea, which is stored and displayed as metadata for the concept (see Fig. 2).

Sample and Data Preparation

We created a simple web crawler to download concepts and inspirations, which exist as individual webpages. The initial sample consisted of 2,344 concepts and 4,557 inspirations posted by 2,452 unique contributors (majority designers, domain experts), for 12 completed challenges; 241 of the concepts are shortlisted by an expert panel for further refinement. In shortlisting, the

Battery collection

Learn from collection initiatives of old batteries.



Battery collection

Consumer can bring old batteries to supermarkets. There collecting containers are installed. Batteries can be recycled and do not pollute the domestic waste.

Since people go to the supermarket anyways the hurdle to overcome is quite low.

The organization GRS was founded in 1998 and is installed in Germany. During 2011 44,7 % of the sold batteries were collected after the usage for recycling via the system. You can find more information at <http://www.grs-batterien.de/> and in English: <http://www.grs-batterien.de/home.html>

GRS is the largest take-back system in Europe.

Figure 3. Example inspiration.

Digital data transfer and elimination service

At the end of a mobile phone's (first) life a data transfer and elimination service will encourage users to hand in their old phones to be re-used, re-furbished or recycled without having to worry that personal data is either lost or misused.



Drawer full of old phones Image from: <http://log.techpayout.com/wp-content/uploads/2013/05/gadget-drawer-1.jpg>

Take-back schemes for electronic gadgets such as mobile phones do exist but despite financial incentives uptake has been slow. One issue is that users have data on devices that they either cannot transfer to a new gadget or the worry that personal data left on the device could be misused by a third party. The proposed digital data transfer and elimination service could be offered by phone retailers. The benefits for retailers include creating a stronger 'green' corporate image, access to resources that can either be re-used or recycled and potential customers looking for a new phone.

Figure 4. Example concept.

expert panel considers both the novelty and feasibility/quality of each concept.

We then created a simple HTML parser to extract the full-text description of each concept/inspiration (for measurement of conceptual distance), and for all concepts, 1) information on which inspirations were cited as sources, and 2) an indicator for whether the concept was shortlisted for development. 707 concepts cited at least one inspiration as a source, with most building on ~10 (median = 10). These 707 concepts (and the 2,826 inspirations that were cited) formed the final sample for analysis

Measuring Conceptual Distance

We used Latent Dirichlet Allocation [8], a form of probabilistic topic modeling, to learn a high-dimensional topic space from the full-text descriptions of the challenge briefs and concepts/inspirations. While most concepts/inspirations included images or video, these media complemented/augmented rather than replaced the text description of the concepts/inspirations. To reduce potential noise, stopwords (e.g., "the", "which") were removed from the text. 750 topics were statistically inferred from the entire collection of 6,913 documents. We then computed the cosine similarity between each inspiration and its challenge brief when projected into the topic space, subtracting this score from 1 so that a higher number would indicate greater conceptual distance. This measure correlated well with human judges' similarity ratings (5 judges, intra-class correlation coefficient = .735) for a subset of the data (199 document pairs, $r = .485$, equal to the highest agreement between the judges). Each concept's "distance" score was the mean distance of its cited inspirations from the challenge brief.

Results

We conducted a logistic regression analysis, using each concept's mean distance of inspirations as the *predictor* variable, and shortlist status as the binary *outcome* variable. The overall model was statistically significant, $\chi^2(1) = 8.42$, $p < .01$, with adequate fit, Hosmer and Lemeshow $\chi^2(8) = 11.32$, $p = .18$ (higher p-value is better fit).

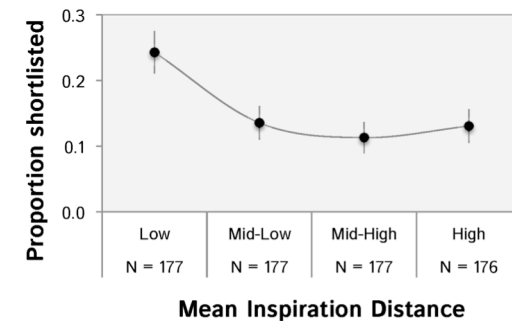


Figure 5. Mean proportion of shortlisted concepts for 4 mean inspiration distance groups (split by quartiles to have equal N), with ± 1 standard error of the mean. Proportion of concepts shortlisted decreases as conceptual distance increases.

The model estimated that a 1-point increase in a concept's mean distance score predicted a *decrease* in its probability of being shortlisted, $\beta = -.31$, Wald (1) = 8.96, $p < .01$. Descriptive statistical analysis of the data (see Fig. 5) suggested that this negative effect of increased distance was most prominent in the change from low (mostly near sources) to mid-low (slightly more near than far sources) mean distance. Replications of the analyses with different topic model

settings yielded very similar results (i.e., no unique benefits of far sources)

Discussion

These surprising results suggest that preferring mostly far sources of inspiration is not most helpful, as claimed in the literature. Rather, citing more near sources than far sources of inspiration seems to be most helpful. Perhaps this relative mix frees up cognitive resources to develop creative concepts through deeper within-category exploration [9], and/or iteration and refinement [10]. Our results might also differ from prior work due to the expert panel's emphasis on both quality and novelty of ideas, not novelty *per se*.

If both novelty and quality are desired, it may be useful to design ways for contributors in crowd-collaborative innovation platforms to interact with prior ideas that are relatively conceptually close to the current problem. Machine learning methods, such as our conceptual distance measure, could be useful foundations for such interfaces. Further research might explore whether these interfaces encourage deeper, more iterative solution exploration, and also potential interactions between distance and novelty/quality tradeoffs.

Acknowledgements

This work is supported in part by an Andrew Mellon fellowship. We are grateful for helpful conversations with Timothy Nokes-Malach and Kevin Ashley.

References

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

[2] Kulkarni, C., Dow, S. P., & Klemmer, S. R. (2012). Early and repeated exposure to examples improves creative work. *Proc. of Cog. Sci* (2012).

[3] Ward, T. B. 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). (1998), Sofia, Bulgaria.

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

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

[6] Enkel, E., & Gassmann, O. Creative imitation: Exploring the case of cross-industry innovation. *R & D Management*, 40 (2010), 256-270.

[7] Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., & Wood, K. The meaning of “near” and “far”: The impact of structuring design databases and the effect of distance of analogy on design output. *Journal of Mech. Design*, 135 (2013), 021007.

[8] Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3 (2003), 993-1022.

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

[10] Dow, S. P., Heddlestone, K., & Klemmer, S. R. The efficacy of prototyping under time constraints. In *Proc. 7th ACM conf. on creativity and cognition*, (2009).