21ST INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED17
21-25 AUGUST 2017, THE UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, CANADA

USING CROWDSOURCING TO PROVIDE ANALOGIES FOR DESIGNER IDEATION IN A COGNITIVE STUDY

Goucher-Lambert, Kosa; Cagan, Jonathan
Carnegie Mellon University, United States of America

Abstract

Analogical reasoning is a prominent method for human creative design. The design research community has invested significant effort into understanding the process of design by analogy, including the impact of different types of analogies on design thinking and solution characteristics. Yet, generating those analogies is a challenge. The present work investigates whether it is possible to obtain useful analogies from individuals with no domain knowledge. To do this, individuals in a crowd workforce were asked to provide solutions for design problems previously explored in the literature. A text mining approach was used to extract commonly used words from these responses, which then served as analogies for problem solvers with design expertise. Finally, 111 participants were recruited for a cognitive study in which they were asked to solve four design problems using some subset of crowd-sourced analogical inspiration. Results indicate that it is feasible to gather impactful analogies from a crowd workforce. The usefulness of analogies at different analogical distances is highly dependent on the problem itself, highlighting the utility of obtaining analogies using the crowd.

Keywords: Crowdsourcing, Design by analogy, Conceptual design, Human behaviour in design

Contact:
Kosa Goucher-Lambert
Carnegie Mellon University
Mechanical Engineering
United States of America
kgoucher@andrew.cmu.edu

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 8: Human Behaviour in Design, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

Analogical reasoning, and more specifically, design-by-analogy is a well-studied and active area of investigation within the design research community (Casakin and Goldschmidt, 1999; Chan et al., 2011; Linsey, Wood, et al., 2008; Moreno et al., 2014). It has often been observed that design practitioners gain inspiration and insight from different domains, which serve to stimulate the formulation of new ideas during the product development process (Markman et al., 2009; Vattam et al., 2010). Furthermore, the use of analogies in design has been studied to gain an understanding of how analogies affect the ideation process as well as design outcomes.

Significant emphasis has been placed on trying to uncover the types of analogies that are most beneficial to stimulating productive design activity. Psychological theory posits that analogical reasoning hinges on the successful mapping of relations between a source and a target domain (Krawczyk et al., 2010). One open problem within design-by-analogy research is where analogies (i.e., the source) are obtained from in the first place. Although researchers can provide specifically tailored analogies for the design problems that they create, if design by analogy is to be used in practice these analogies have to be systematically available and appropriate for new problems when they are being solved. Initial work in this area included the use of the US patent database to map and identify near to far analogies (Fu et al., 2013), as well as semantic verb mapping (Linsey et al., 2012). This work introduces a different way to identify analogies from individuals with no problem solving or domain expertise. Leveraging the vast power of the crowdsourcing workforce, it is possible to gain access to a high volume of workers for directed tasks. Here, a crowd-based workforce is used to generate analogies for future design problem solvers.

Crowdsourcing describes a model in which a distributed network of individuals respond to an open call for proposals or work (Brabham, 2008; Howe, 2006). There are examples of specialized crowdsourcing design platforms, such as the OpenIDEO (Lakhani et al., 2012), and Local Motors (Norton and Dann, 2011), which offer domain experts the opportunity to create and collaborate with others. However, these platforms seek to use the crowd workforce for both the creation and validation of concepts. Using these services, crowd workers create new solutions, but also vote on designs or ideas that they think are best. These platforms, while practical for rapid innovation in industry, have not been utilized widely in academic research.

To date, the overall use of crowdsourcing within the design research community has been limited. Primarily, the design research community has used crowdsourcing for evaluative and rating purposes, with little to no expression of creativity and self directed input (Kittur et al., 2008; Kudrowitz and Wallace, 2013). For example, Kurowitz and Wallace (2013) used a crowdsourced population to rate design concepts on a number of subjective measures, including creativity, novelty, clarity, and usefulness. One of the reasons for the limited number of crowdsourcing applications in design research literature is that it is difficult to identify individuals with expertise and domain specific knowledge within crowd-based communities. In situations where it has been possible to identify members of the crowd with domain specific knowledge, these individuals have been unable to provide consistent and accurate responses (Burnap et al., 2015). Previous work has demonstrated that online crowdsourcing services, such as Amazon Mechanical Turk (MTurk), provides a participant population pool that is representative of the United States population (Paolacci et al., 2010). In this work, we attempt to leverage an open crowd-based workforce using MTurk for a creative task, in which each individual participating is asked to come up with a solution/idea to an open-ended design problem. For the purposes of this experiment, we find the MTurk population to be appropriate because we assume crowd participants have no level of domain expertise and because the research team has no requirements for “right” or “wrong” answers. The responses from each crowd participant are then text-mined with the purpose of extracting analogical stimuli. We theorize that commonly used words will represent “near” analogies, and infrequently used words will represent “far” analogies.

One of the key factors influencing the process of retrieving relevant information from a source, and then applying useful connections to a target, rests upon the analogical *distance* between the two. Primarily, research on analogical distance uses the terms “near” and “far” to discuss the distance of the analogy from the problem being examined (Fu et al., 2013). The continuum of distance refers to the domain distance; a “near” analogy generally implies that the analogy comes from the same or closely related domain, whereas a “far” analogy comes from a distant domain. It has also been noted that near-field analogies share significant surface level (object) features with the target, and far-field analogies share little or no surface features (Linsey et al., 2012). For example, when trying to design a device to reduce home energy use, a design team could be to take inspiration from smart thermostats, which learn and adjust heating and cooling schedules to match behavior and save energy (near analogy). Another approach could take inspiration from nature in which grazing animals sync their foraging cycles to match plant growth cycles (far analogy).

That said, it is not clear which analogical distance is most likely to yield positive solution characteristics in a given problem. Several studies support the idea that more distant analogies positively impact ideation (Ward, 1998; Wilson et al., 2010). There are several anecdotal examples regarding the use of analogy in design and product development. A review regarding the use of analogies in industry found that far-field analogies are more beneficial in helping to create more novel solutions (Kalogerakis et al., 2010). However, some empirical evidence disputes this (Chan et al., 2015). Fu et al. (2013) proposed that there exists a "sweet spot" of analogical distance that rests between an analogy being too near (where innovation is restricted and fixation and copying are likely to occur) and too far. This work further contributes to this discussion by examining the differences in solution characteristics that are observed when the distance of the analogical stimuli is varied. By classifying the crowdsourced data into near, medium and far categories, any difference in impact can be assessed. Using these generated analogies, we examine their impact on several solution characteristics (e.g., novelty and quality) for concepts generated for multiple design problems by human problem solvers with some level of domain expertise.

The work presented in this paper looks to combine and explore crowdsourcing with design by analogy. In particular, there are two main aims of this paper. First, we seek to determine whether or not crowdsourcing is a viable means to obtain meaningful analogical stimuli for future design problem solvers. Using the extracted analogies obtained from the crowdsourced design solutions, a continuum of analogical distance is created. Second, we aim to use analogies from the crowd at varying levels of distance to examine the impact of analogical distance on solution characteristics.

2 METHODS

Figure 1 shows the four step methodological outline governing the experiment. First, twelve conceptual design problems were identified from the design research literature. Next, these twelve problems were posted online in an open call for crowd responses. With over 1000 responses obtained between the 12 problems, the textual data was examined using a natural language processing toolkit. Commonly used words were extracted as analogies for a human subject study performed using a subset of 4 of the original 12 design problems. Three experimental conditions were explored, each of which varied the distance of the analogical stimuli from the problem statement. Results were analysed to determine the impact of the analogies on the quantity, quality, and novelty of solutions generated by the human subject participants.

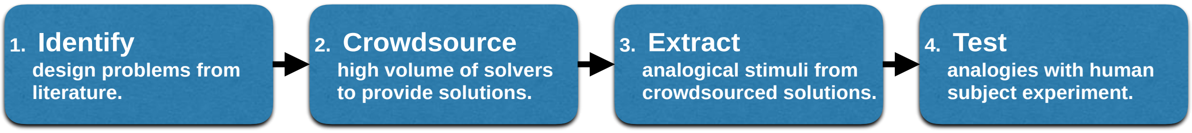

Figure 1. Methodological outline of experiment

2.1 Selecting Design Problems

Through a review of the design-by-analogy literature, 12 design problems used in prior research were chosen subjectively by the authors to include in the study. With the knowledge that these problems would be used within a crowdsourcing environment, some of them were modified such that design constraints were removed. This was done primarily to limit the required time to provide a single idea for the problem to a few minutes, and to allow the crowd population (with no design domain expertise) to successfully provide a relevant idea. A diversity of problem domains was also sought in selecting the design problems. The adapted versions of each design problem used within the current study and relevant references are shown in Table 1. The modified forms of the problems were limited to a single sentence. Problem 13 (Table 1) was developed by the authors.

Table 1. Design problems selected from literature for crowdsourcing experiment

Problem	Reference
1. A lightweight exercise device that can be used while traveling.	(Linsey and Viswanathan, 2014)
2. A device that can collect energy from human motion.	(Fu et al., 2013)
3. A new way to measure the passage of time.	(Tseng et al., 2008)
4. A device that disperses a light coating of a powdered substance over a surface.	(Linsey et al., 2008)
5. A device that allows people to get a book that is out of reach.	(Cardoso and Badke-Schaub, 2011)
6. An innovative product to froth milk.	(Toh and Miller, 2014)
7. A way to minimize accidents from people walking and texting on a cell phone.	(Miller et al., 2014)
8. A device to fold washcloths, hand towels, and small bath towels.	(Linsey, et al., 2008)
9. A way to make drinking fountains accessible for all people.	(Goldschmidt and Smolkov, 2006)
10. A measuring cup for the blind.	(Jansson and Smith, 1991)
11. A device to immobilize a human joint.	(Wilson et al., 2010)
12. A device to remove the shell from a peanut in areas with no electricity.	(Viswanathan and Linsey, 2013)
13. A device that can help a home conserve energy.	N/A

2.2 Crowdsourcing Design Problem Solutions

The 13 design problems shown in Table 1 were posted on MTurk, an online crowdsourcing labour market. Each problem was posted as a separate Human Intelligence Task (HIT), where the requesters (in this case, the authors) sought 100 responses from unique workers for each design problem. In total, 1345 responses were made for the HITs. There were 45 rejected submissions due to workers not submitting fully completed assignments. The 97% acceptance rate for the HITs as a part of this work is in line with other MTurk submissions, as workers in the crowd-based community desire a high approval rating to garner more HIT opportunities. Workers responded to each HIT in return for $0.20. No time limit was placed on the task and no demographic information was sought through the collection of data. The only requirements placed through MTurk were that all workers were required to be US citizens and at least 18 years of age.

For each HIT, workers were asked to provide an idea (solution) for a new product or device that addressed the given prompt. The instructions for the HIT asked that the provided idea be something that workers believed did not currently exist. Workers were also instructed that they should not be concerned how, or if, what they were thinking of would be made. Once workers thought of an idea, they were asked to use as many words as necessary to describe it by writing into a free response text box. Next, participants were asked to provide up to six keywords (three nouns, three verbs) to serve as identifiers for the idea that they had entered into the free response box. Although only the keywords were of interest for this study, initial research showed that participants were more likely to provide accurate keywords if they pertained to an idea that they had already thought of.

2.3 Extracting Analogical Stimuli

The three noun and three verb keywords provided with each HIT response from the MTurk task were used to obtain design analogies at varying distances. Word frequency was used as a measure of distance. Commonly used words within the response set were taken as near field analogies, and infrequently used words were taken as far analogies. Due to the fact that word frequency provided a continuous distribution of words, a "medium" distance field set was also extracted from the crowd-sourced responses. To accomplish this, the raw text from MTurk HIT responses was first collated together for each design problem. Using Python's Natural Language Processing Toolkit, individual word tokens were extracted from the raw text (Bird and Loper, 2004). The word token set was cleaned by removing stop words (e.g. "the", "is", "that", etc.), words that appeared in the problem statement (e.g. "reach" from Problem 5, "A device that allows people to get a book that is out of reach"), and by aggregating multiple tenses of words (e.g. "reach", "reaching", etc.). Following this, the new cleaned token set were used to create a frequency distribution of words. The words that accounted for the top 25% of the frequency distribution were pooled into the near analogy group, words used multiple times after 30% were used as the medium set, and words used only once became the far set.

2.4 Exploring Analogical Distance Using a Human Subject Cognitive Study

2.4.1 Participants

Participants for the cognitive study were recruited from upper level design and innovation courses at Carnegie Mellon University and offered course credit or $10 compensation for their participation. 95 participants were recruited from junior and senior level mechanical engineering design courses. An additional 16 participants were recruited from a multidisciplinary course focusing on design innovation. There were 67 male and 44 female participants ranging in age from 19 to 26.

2.4.2 Conditions

Four conditions (three experimental, one control) were explored using the crowdsourced analogical stimuli. These conditions varied the distance of the analogy from the problem, defined in this experiment as the word frequency from the text-mined crowdsourced data. Analogical stimuli for the near, medium, and far conditions were extracted using the methods outlined in Section 2.3. Each of the four conditions randomly assigned four words from within the available word set (approximately 600 words classified into three sets per problem). The control condition re-displayed four words from the problem statement, which were assumed to provide no additional analogical stimulus to participants.

2.4.3 Cognitive Study Procedure

The cognitive study involved an approximately 1-hour session during which participants were asked to “come up with concept solutions to open ended design problems”. Participants were told that they might receive a set of words during the problem that were intended to serve as inspiration for their concepts. Each participant saw the same four of the original thirteen design problems (4, 7, 12, and 13 from Table 1) used in the crowdsourcing experiment. These four design problems were selected for the cognitive study due to high lexical diversity in their solution word set from the crowdsourced data and low completion time. A full factorial experimental design evenly split the conditions for each problem across study groups (Table 2), such that a given participant only saw one of the four conditions (near analogy, medium analogy, far analogy, or control) for a given design problem.

Table 2. Cognitive study group conditions

Problem	Group A (N= 28)	Group B (N=28)	Group C (N=29)	Group D (N=26)
4	Near	Medium	Far	Control
7	Medium	Far	Control	Near
11	Far	Control	Near	Medium
12	Control	Near	Medium	Far

At the start of the experiment, participants were provided with envelopes containing four separately marked problem packets, each containing a separate design problem. Participants were given ten minutes to work on each design problem, divided into two working blocks. Participants began each problem by first spending two minutes working to provide a single solution, along with up to six descriptors (three nouns, and three verbs) for the design problem. This initial procedure was meant to mirror the crowdsourced data; however, the descriptors generated by the cognitive study participants were not used to generate analogies. One reason for having participants initiate each brainstorming session before receiving the crowd-sourced analogies, is that prior research on analogical reasoning has shown that analogies are more effectively applied if an open goal has been established (Tseng et al., 2008). After this initial period, participants were provided the crowd-sourced analogical stimuli specific to their condition. These stimuli consisted of 4 words extracted from the text-mined MTurk dataset. Eight minutes of open idea generation was given following the presentation of the analogies, where participants placed each generated idea into individual designated boxes provided within the problem packet. Each idea was time stamped at completion by the participant using a clock displayed at the front of the room during the study. Participants were allowed to use any combination of sketching and writing to express their ideas, and were instructed to provide sufficient detail such that someone viewing their ideas later could understand the basic concept. Following each problem, a short questionnaire was provided to gauge participant's perceived usefulness and relevancy of the presented analogies, as well as the self-perceived quality and novelty of their generated solutions.

2.4.4 Analysis

The design output from the participants was examined in order to determine the impact of crowdsourced analogical stimuli at varying distances on solution characteristics. The following characteristics of the solution outputs were explored:

1. Quantity: the number of unique ideas generated for a given problem.
2. Feasibility: rated on an anchored scale from 0 (the technology does not exist to create the solution) to 2 (the solution can be implemented in the manner suggested).
3. Novelty: rated on an anchored scale from 0 (the concept is copied from a common and/or pre-existing solution) to 2 (the solution is new and unique).
4. Usefulness: rated on an anchored scale from 0 (the solution does not address the prompt and/or take into account implicit problem constraints) to 2 (the solution is helpful beyond status quo).
5. Quality: rated subjectively by each rater on a scale from 0 (low) to 2 (high).

In addition, perceived attributes for novelty and quality were extracted from the participants (rated 1-5), along with the perceived usefulness and relevancy of the provided analogies.

One mechanical engineering PhD student and one mechanical engineering postdoctoral researcher, both specializing in design theory and methodology, were trained to perform solution characteristic ratings. Consistency was assessed over a subsample of the data using the intraclass correlation coefficient (ICC) (Shrout and Fleiss, 1979).

3 RESULTS

3.1 Crowdsourced Analogies

Using the methods outlined in Section 2.3, analogies were extracted from the crowdsourced dataset provided by 1345 respondents, using word commonness as a measure of analogical distance. Five analogies were extracted for each distance measure, and are shown in Table 3. Table 3 also shows the average response time for the HIT and the lexical diversity of the solution for each set of analogies. The average completion time provided insight into the difficulty of the problems for the crowd community. Lexical diversity measures the ratio of unique word entries within the submissions. Both of these measures were used to select four problems from the available thirteen to use in the cognitive study, with a low response time and high lexical diversity seen as positive problem characteristics. The design problems selected for the cognitive study were Problems 4, 7, 11, and 12. For brevity, only these four problems are included in Table 3.

Table 3. Extracted Analogies, Solution Time, and Lexical Diversity of solutions from crowd-sourced concept generation experiment

Problem	Avg. Time (s)	Lexical diversity	Near Words	Middle Words	Far Words
4	199	0.539	spray, blow, fan, shake, squeeze	handle, puff, mesh, reservoir, hand	rotor, wave, cone, pressure, atomizer
7	207	0.530	alert, flash, camera, sensor, motion	smart, beep, notify, background, recognize	emit, react, engage, lens, reflection
11	179	0.636	splint, wrap, hold, harden, apply	lever, spray, strap, slide, magnet, inflate	shrink, inhale,fabric, condense, pressure
12	194	0.501	crack, crank, blade, squeeze, conveyor	pry, spin, mill, fall, drop	melt, circular, wedge, chute, wrap

3.2 Cognitive Study Results

111 participants generated 1651 concepts across the four design problems. Each solution was rated using the methods outlined in Section 2.4.4. In addition to the rated solution characteristics, participants also

provided perceived rating values for the relevancy and usefulness of the presented analogical stimuli, as well as the quality and novelty of their own solutions.

Both raters evaluated a randomly selected subset of 150 solutions across the sub-dimensions of interest (Usefulness, Feasibility, Novelty, Quality) and consistency was assessed using the intraclass correlation coefficient (ICC). A strong level of correlation was obtained for three of the four metrics: Usefulness (ICC >0.65), Novelty (ICC > 0.71), and Feasibility (ICC > 0.77). ICC for the Quality metric was fair at ICC > 0.50. The inter-rater reliability levels for this study are within the range of values typically found in behavioural studies with human raters (Cicchetti, 1994).

3.2.1 Aggregated Solution Characteristics

Results were analysed across the four design problems for each condition of interest (near, medium, far analogical distance, and control). Figure 2 (Aggregated Problems) displays the mean and standard error for each metric, with pooled data from all four problems. In addition, four separate one-way ANOVAs were run for each condition. A weakly significant effect of analogical stimuli against the control condition was found for Feasibility ($F(3,1609)= 2.07$, $p = 0.10$), and no significant effects were found for the other three metrics (Usefulness, ($F(3,1609)= 1.28$, $p = 0.28$), Novelty, ($F(3,1609)= 1.46$, $p = 0.220$), and Quality, ($F(3,1609)= 1.24$, $p = 0.29$)). For all metrics, the highest mean value was found in one of the three analogical distance conditions. However, when aggregated across design problems, this difference was largely found to not be statistically significant.

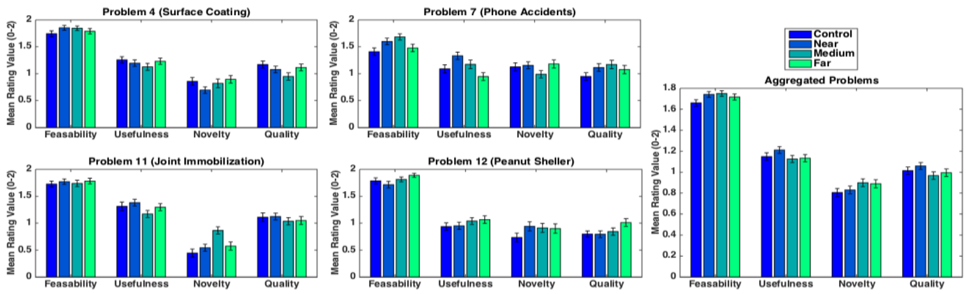

Figure 2. Mean rating values (+/- 1 SE) for cognitive study design solution characteristics for each design problem, and aggregated across all four problems

Table 4. Solution characteristics and overall impressions specific to each design problem

Problem	Measure	Feasability	Usefulness	Novelty	Quality
4 (Surface Coating)	**One-way ANOVA all conditions**	$F(3, 380) = 1.12$, $p = 0.34$	$F(3, 380) = 0.77$, $p = 0.50$	$F(3, 380) = 1.57$, $p = 0.19$	$F(3, 380) = 1.95$, $p = 0.121$
	One-way ANOVA analogy conditions	N/A	N/A	$F(2, 275) = 2.21$, $p = 0.11$	N/A
	Impressions	No significant difference	No significant difference	Far analogies are slightly better than near analogies (*p=0.094, 95% CI).*	No significant difference
7 (Phone Accidents)	**One-way ANOVA all conditions**	$F(3, 392) = 3.69$, $p = 0.01$	$F(3, 392) = 4.45$, $p = 0.004$	$F(3, 392) = 1.41$, $p = 0.24$	$F(3, 392) = 6.30$, $p = 0.0003$
	One-way ANOVA analogy conditions	$F(2, 282) = 2.77$, $p = 0.06$	$F(2, 282) = 6.79$, $p = 0.001$	$F(2, 282) = 2.28$, $p = 0.10$	$F(2, 282) = 8.99$, $p = 0.0002$
	Impressions	All analogies improve Feasability compared to control. Medium distance analogies best (*p = 0.05, 95% CI).*	Near analogies improve usefulness. Analogical distance negatively impacts usefulness-- Near analogies much better than far ($p = 0.0009$, 95% CI).	Analogies don't improve novelty compared to control.	Near analogies are most beneficial (*p = 0.002, 95% CI),* but all analogies improve quality against control.
11 (Joint Immobilization)	**One-way ANOVA all conditions**	$F(3, 416) = 0.22$, $p = 0.88$	$F(3, 416) = 1.74$, $p = 0.15$	$F(3, 416) = 7.22$, $p = 0.001$	$F(3, 416) = 0.38$, $p = 0.75$
	One-way ANOVA analogy conditions	N/A	$F(2, 317) = 2.75$, $p = 0.06$	$F(2, 317) = 6.71$, $p = 0.001$	N/A
	Impressions	No significant difference	Near analogies are best (*p = 0.05, 95% CI).* Far analogies are also good, but medium do not help.	Medium analogies are best (*p < 0.01, 95% CI).* Near and far analogies are about equal (*p = 0.95, 95% CI).*	No significant difference
12 (Peanut Sheller)	**One-way ANOVA all conditions**	$F(3, 388) = 1.94$, $p = 0.12$	$F(3, 388) = 0.89$, $p = 0.44$	$F(3, 388) = 1.23$, $p = 0.29$	$F(3, 388) = 2.30$, $p = 0.07$
	One-way ANOVA analogy conditions	$F(2, 294) = 2.97$, $p = 0.05$	N/A	N/A	$F(2, 294) = 2.70$, $p = 0.06$
	Impressions	Far analogies are best (*p = 0.04, 95% CI).*	No significant difference	All analogies improve novelty compared to control (*p= 0.07, 95% CI).*	Far analogies lead to highest quality (*p = 0.06, 95% CI)*

3.2.2 Solution Characteristics Separated by Design Problem

To gain more insight into the effect of varying analogical distances on solution characteristics, results were analysed on an individual problem basis. Figure 2 (Problem 4, 7, 11, 12) displays the mean and standard error for all solution characteristics, analysed separately for each of the four design problems. Additionally, one-way ANOVAs were conducted with all four experimental conditions for each specific problem. In some situations (or if the results from the initial one-way ANOVA showed significant differences) a separate one-way ANOVA was performed with the three analogical distance conditions only. If one-way ANOVAs showed significant differences in the mean values the condition contrasts were further analysed using the Tukey HSD (honest significant difference) test to obtain significance values at a 95% confidence interval. Key results from these analyses are included in Figure 2, along with overall interpretations from the analyses.

3.2.3 Participant Provided Ratings

Participants provided four ratings following the presentation of each problem. Two of these were gauged at assessing the analogical stimuli that were presented for each design problem (usefulness and relevance), and the other two sought to determine participants' subjective perception regarding the overall novelty and quality of the solutions they developed for that problem (Figure 3). There was a clear trend in perceived usefulness and relevancy of analogical stimuli, where study participants perceived less distant analogies to be more useful and more relevant (Usefulness, $F(2,309) = 3.74, p = 0.025$; Relevancy, $F(2,309) = 18.26, p < 0.001$). However, there was no significant difference between how participants perceived the quality or novelty of their own solutions within the different conditions (Quality, $F(3,412) = 0.73, p = 0.53$; Novelty, $F(3,412) = 1.25, p = 0.29$). The data trend indicates that participants perceived their solutions to be more novel as analogical distance increased.

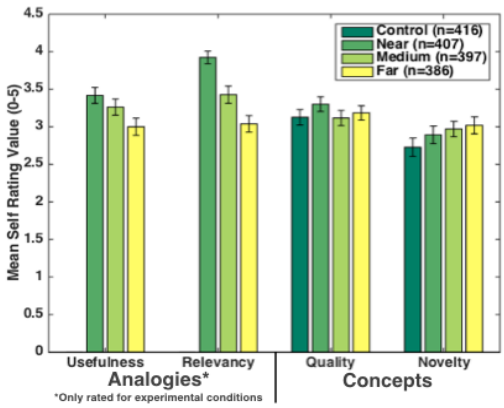

Figure 3. Participant provided ratings of analogical stimuli and generated concepts (+/-1 SE)

4 DISCUSSION

This work uses crowdsourcing to obtain analogical stimuli for future design problem solvers. By text-mining design solutions from crowd participants with no design expertise, commonly used words are able to be extracted, and later serve as analogies. Here, more common words were used to specify "near" analogies, and less commonly used words serve as "far" analogies. A cognitive study tested these analogies on participants with design domain expertise (all participants were students currently enrolled in undergraduate level design courses), as they solved four conceptual design problems.

Results indicate that the methods employed in this work for crowdsourcing analogical stimuli, and using word commonness as a measure of analogical distance, were successful. Participants in the cognitive study rated more distant analogies as having a lower level of relevancy to the design problem (Figure 3). This is in line with previous research regarding analogical distance and participant perception of the relevance of analogies to the problem domain (Fu et al., 2013). The drop in relevancy between experimental conditions (near to medium to far) shows that this effect is robust across multiple levels of analogical distance. It is also interesting to note that participants rated more distant analogies as being less useful. This indicates that participants likely had difficulty connecting distant analogies to the design problem domains. Further evidence of the difficulty participants had in applying more distant analogies to the problem domain can be seen when examining the number of solutions generated in each condition.

Across all problems, the number of solutions decreased along with increasing analogical distance (Control = 416, Near = 407, Medium = 397, Far = 386).

However, participants are not always able to successfully discern how they apply an analogy, or whether or not that analogy is useful. A separate goal of this work was to link the distance of analogical stimuli to a variety of solution characteristics. The four design problems included in the cognitive study came from various domains. When aggregating the results across design problems for each condition, there was only a small effect of analogical distance (Figure 2). When aggregated, none of the four solution characteristics had statistically significant mean values across experimental conditions. However, a deeper examination into the solutions developed for each individual problem indicates that the analogical stimuli greatly impacted solutions for each of the design problems (Table 4). This demonstrates that the effect of analogical distance appears to be intimately linked to properties and/or the domain of the problem itself, consistent with the "sweet spot" proposed by Fu et al. (2013). For example, in Problem 12 (Peanut Sheller) more distant analogies led to more feasible and higher quality solutions, where as in Problem 7 (Phone Accidents) near analogies helped participants develop more useful and higher quality concepts. Additional work is needed to develop and test theories regarding the specific problem properties that are better suited for a specific analogical distance. Additionally, it is possible that the *effective* distance of the analogical stimuli may have varied for each problem (i.e. some "far" words might not have been far).

Not only do the results of this work indicate that the effectiveness of an analogy is dependent on the specific problem, but also that it is unclear what type of analogy will be most beneficial to promote positive solution characteristics. With this in mind, obtaining a large, diverse, and continuous set of analogies for a given problem is beneficial. Using a crowd workforce, this can be accomplished quickly and effectively, as demonstrated in this work. Future research can investigate ways to further automate this process for the quality extraction of analogical stimuli.

5 CONCLUSIONS

This work examined whether it is feasible to obtain analogical stimuli using crowdsourcing techniques and how these sourced analogies impact solution characteristics of design concepts generated by participants in a cognitive study. Results indicate that it is possible to obtain analogical stimuli effectively using an untrained crowd workforce. Furthermore, the analogies from crowdsourced design solutions are able to effectively translated onto a continuous space of analogical distance. When testing the impact of analogical distance of the crowdsourced analogies on solution characteristics, we find that aggregating results across different problems negates the significance of the effect. Instead, results separated by individual problems are significant. This indicates that the effectiveness of solutions at varying analogical distances is highly dependent on the problem itself. Additional work is needed to fully understand when problem solvers will benefit from having analogical stimuli.

REFERENCES

Bird, S. and Loper, E. (2004), "NLTK: The Natural Language Toolkit", *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pp. 1–4.

Brabham, D.C. (2008), "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases", *Convergence: The International Journal of Research into New Media Tech.*, Vol. 14 No. 1, pp. 75–90.

Burnap, A., Ren, Y., Gerth, R., Papazoglou, G., Gonzalez, R. and Papalambros, P.Y. (2015), "When Crowdsourcing Fails: A Study of Expertise on Crowdsourced Design Evaluation", *Journal of Mechanical Design*, Vol. 137 No. 3, p. 031101.

Cardoso, C. and Badke-Schaub, P. (2011), "The influence of different pictorial representations during idea generation", *Journal of Creative Behavior*, Vol. 45 No. 2, pp. 130–146.

Casakin, H. and Goldschmidt, G. (1999), "Expertise and the use of visual analogy: implications for design education", *Design Studies*, Vol. 20 No. 2, pp. 153–175.

Chan, J., Dow, S.P. and Schunn, C.D. (2015), "Do the best design ideas (really) come from conceptually distant sources of inspiration?", *Design Studies*, Elsevier Ltd, Vol. 36 No. C, pp. 31–58.

Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K. and Kotovsky, K. (2011), "On the Benefits and Pitfalls of Analogies for Innovative Design", *Journal of Mechanical Design*, Vol. 133 No. 8, p. 081004.

Cicchetti, D. V. (1994), "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology", *Psychological Assessment*, Vol. 6 No. 4, pp. 284–290.

Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C. and Wood, K. (2013), "The Meaning of 'Near' and 'Far'", *Journal of Mechanical Design*, Vol. 135 No. 2, p. 021007.

Goldschmidt, G. and Smolkov, M. (2006), "Variances in the impact of visual stimuli on design problem solving performance", *Design Studies*, Vol. 27 No. 5, pp. 549–569.
Howe, J. (2006), "The Rise of Crowdsourcing by Jeff Howe | Byliner", *Wired Magazine*.
Jansson, D.G. and Smith, S.M. (1991), "Design fixation", *Design Studies*, Vol. 12 No. 1, pp. 3–11.
Kalogerakis, K., Lüthje, C. and Herstatt, C. (2010), "Developing innovations based on analogies: Experience from design and engineering consultants", *J. of Product Innovation Mgmt*, Vol. 27 No. 3, pp. 418–436.
Kittur, A., Chi, E.H. and Suh, B. (2008), "Crowdsourcing user studies with Mechanical Turk", *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems*, p. 453.
Krawczyk, D.C., McClelland, M.M., Donovan, C.M., Tillman, G.D. and Maguire, M.J. (2010), "An fMRI investigation of cognitive stages in reasoning by analogy.", *Brain Research*, Vol. 1342, pp. 63–73.
Kudrowitz, B.M. and Wallace, D. (2013), "Assessing the quality of ideas from prolific, early-stage product ideation", *Journal of Engineering Design*, Vol. 24 No. 2, pp. 120–139.
Lakhani, K., Fayard, A.-L., Levina, N. and Pokrywa, S.H. (2012), *OpenIDEO*, *New York*, Vol. N1–612–066
Linsey, J., Markman, A.B. and Wood, K.L. (2008), "WordTrees: A method for design-by-analogy", *Proceedings of the 2008 ASEE Annual Conference*.
Linsey, J.S., Markman, a. B. and Wood, K.L. (2012), "Design by Analogy: A Study of the WordTree Method for Problem Re-Representation", *Journal of Mechanical Design*, Vol. 134 No. 4, pp. 041009–041009.
Linsey, J.S. and Viswanathan, V.K. (2014), "Overcoming Cognitive Challenges in Bioinspired Design and Analogy", *Biologically Inspired Design*, pp. 221–244.
Linsey, J.S., Wood, K.L. and Markman, a. B. (2008), "Modality and representation in analogy", *Ai Edam*, Vol. 22 No. 02, pp. 85–100.
Markman, A.B., Wood, K.L., Linsey, J.S., Murphy, J.T. and Laux, J.P. (2009), "Supporting Innovation by Promoting Analogical Reasoning", *Tools for Innovation*, available at:https://doi.org/10.1093/acprof:oso
Miller, S.R., Bailey, B.P. and Kirlik, A. (2014), "Exploring the utility of Bayesian Truth Serum for assessing design knowledge.", *Human-Computer Interaction*, Vol. 29 No. 5-6, pp. 487–515.
Moreno, D.P., Hernandez, A. a., Yang, M.C., Otto, K.N., Holtta-Otto, K., Linsey, J.S., Wood, K.L., et al. (2014), "Fundamental studies in Design-by-Analogy: A focus on domain-knowledge experts and applications to transactional design problems", *Design Studies*, Vol. 35 No. 3, pp. 232–272.
Norton, M. and Dann, J. (2011), "Local Motors : Designed by the Crowd , Built by the Customer", *Harvard Business Review*, Vol. 9-510-062 No. September, pp. 1–21.
Paolacci, G., Chandler, J. and Ipeirotis, P.G. (2010), "Running experiments on amazon mechanical turk", *Judgment and Decision Making*, Vol. 5 No. 5, pp. 411–419.
Purcell, a T., Williams, P., Gero, J.S. and Colbron, B. (1993), "Fixation effects: do they exist in design problem solving?", *Environment and Planning B: Planning and Design*, Vol. 20 No. 3, pp. 333–345.
Shrout, P.E. and Fleiss, J.L. (1979), "Intraclass correlations: Uses in assessing rater reliability.", *Psychological Bulletin*, Vol. 86 No. 2, pp. 420–428.
Toh, C. a and Miller, S.R. (2014), "The Impact of Example Modality and Physical Interactions on Design Creativity", *Journal of Mechanical Design (Transactions of the ASME)*, Vol. 136 No. 9, p. [np].
Tseng, I., Moss, J., Cagan, J. and Kotovsky, K. (2008), "The role of timing and analogical similarity in the stimulation of idea generation in design", *Design Studies*, Vol. 29 No. 3, pp. 203–221.
Vattam, S.S., Helms, M.E. and Goel, A.K. (2010), "A content account of creative analogies in biologically inspired design", *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Carnegie Mellon University, Vol. 24 No. 04, pp. 467–481.
Viswanathan, V.K. and Linsey, J.S. (2013), "Design Fixation and Its Mitigation: A Study on the Role of Expertise", *Journal of Mechanical Design*, Vol. 135, p. 051008.
Ward, T.B. (1998), "Analogical distance and purpose in creative thought: mental leaps versus mental hops", in K. Holyoak, D. Gentner, B.K. (Ed.), *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*, New Bulgarian University Press, pp. 221–230.
Wilson, J.O., Rosen, D., Nelson, B. a. and Yen, J. (2010), "The effects of biological examples in idea generation", *Design Studies*, Elsevier Ltd, Vol. 31 No. 2, pp. 169–186.

ACKNOWLEDGMENTS

The authors would like to thank Josh Gyory and Chris McComb for their help rating design solutions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under grant DGE125252, the United States Air Force Office of Scientific Research through grant FA9550-16-1-0049, and the CIT Bradford and Diane Smith Fellowship at Carnegie Mellon University. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.