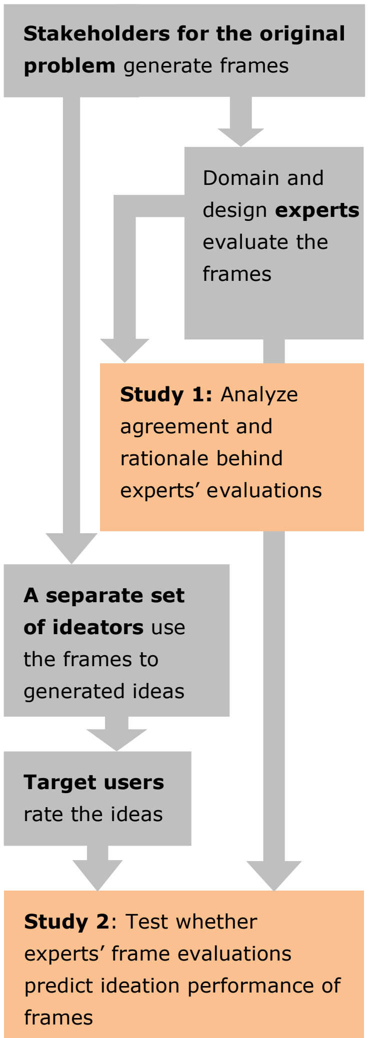
An Exploratory Study of Problem Framing in Distributed Collaborative Design

Figure 1: Overview of our study



Ruijia Cheng
University of California, San Diego, CA 92093, USA
Ruc019@ucsd.edu

Jan De Castro
University of California, San Diego, CA 92093, USA
Jedecast@cs.cmu.edu

Steven Dow
University of California, San Diego, CA 92093, USA
Spdow@ucsd.edu

Joel Chan
Carnegie Mellon University, Pittsburgh, PA 15213, USA
Joelchuc@cs.cmu.edu

Abstract
Design is increasingly conducted in distributed, online, and asynchronous settings (such as online ideation platforms). Effectively framing (the goals, constraints, possible solutions of) a problem is critical for effective design, but challenging in these settings. We explore whether and how expert knowledge can be leveraged to help distributed design teams select effective frames for design problems. We show that experts with different knowledge bases (design, and domain expertise) have reliably quantified opinions about what makes for effective frames, and that these opinions are useful for predicting high/low performing frames with respect to novelty, focus, and quantity of ideas generated with those frames. These results suggest that distributed design teams could benefit from structured problem framing processes that incorporate domain and design experts.

GROUP '18, Jan 7 - 10 2018, Sanibel Island, Florida USA

Author Keywords
Problem framing; distributed collaborative design

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

Introduction
Framing is a critical part of the design process, which defines the goals, constraints and possible solutions of the problem being solved [1,2,3]. Design today is increasingly being done in *distributed* collaborative settings, such as OpenIDEO.com, an online platform where expert designers frame problems that are solved by crowds of volunteer novice designers. The distributed and asynchronous nature of interactions on these platforms makes it difficult and inefficient to effectively frame using typical strategies like back-and-forth between designers offering frames and gauging resonance with teammates [1], or trying out solutions and refining frames [2]. How can distributed design teams effectively frame problems?

Table 1. Example frames with overall ratings. Quality scale is from -3 (worst) to +3 (best)

Sample "good" frames:
- What are ways that voting is currently inaccessible to young voters, and how can we reduce these barriers? <u>(rating = 1.73)</u>
- How should we use the internet to motivate younger generation to vote? <u>(rating = 0.8)</u>
- How might we increase youth voter turnout through the impact of the parents onto their children? <u>(rating = 0.6)</u>
Sample "bad" frames:
- How might we convince employed young people to vote? <u>(rating = -0.13)</u>
- How does the government market young adults to vote? <u>(rating = -0.27)</u>
- Based on current states is there a correlation between parents' education/interest and youth votes? <u>(rating = -1.27)</u>

In this paper, we explore whether and how expert knowledge (in the problem domain or in design process) can help distributed design teams select effective problem frames. Specifically, we explore two questions: 1) Can we reliably quantify experts' opinions about what makes for effective frames? and 2) (how) do those quantified opinions predict which frames will do better/worse when they are actually used by others for idea generation?

Figure 1 shows an overview of our approach to exploring these questions. Across two studies with domain experts, designers, and online crowd ideators collaborating in a distributed fashion for a real-world design problem (increasing voter turnout among young people), we find that experts' opinions about (in)effective frames can be reliably quantified, and that domain experts and design experts bring complementary useful perspectives for selecting good frames: domain experts' ratings can predict frames that will lead to less novel ideas, and select frames that will help focus ideation, while design experts' ratings can predict frames that are generative and lead to higher quantities of ideas.

These results suggest practical strategies for effective problem framing during distributed collaborative design, and suggest that it may be possible to systematize the process of effectively framing design problems.

Task/design context and frames

We explore the framing process in the context of a real-world problem: *"What is a more efficient way in getting the younger generation to come out and vote (increase voter turnout)?"* This problem is concrete, yet broad enough to allow a variety of frame perspectives.

In order to get a broad and diverse coverage of the frame space, we recruited 9 participants representing a range of relevant stakeholders for this topic: youths (ages between 18-20 years old), educators, civic engagement activists and researchers, and general designers. Participants completed an online survey which first reviewed the concept of framing and then asked them to generate as many different frames as possible for the youth voting problem.

Participants generated a total of 40 unique frames across a range of themes, including "publicity/technology", "school/education", "accessibility", "parental/family actors", "incentive" and "influence/encouragement".

Study 1: Can Experts Opinions About What Makes for Effective Frames be Reliably Quantified?

Method

Recruiting appropriate experts

While there are many possible aspects to what makes for good/bad frames, some of them might be specific to the problem (e.g., what specific aspect of the problem is emphasized), and others general across problems (e.g., how the frame is described). In order to get a throughout understanding for what makes for good/bad frames from both aspects, we recruited professors, professionals, postdoc or Ph.D students in the fields of design, political science, and civic engagement. Experts completed a brief screening survey, self-rating their experience in the fields of design and the youth voting problem (separately), on a scale of 1 (almost no experience) to 4 (at least professional-level experience). We then selected experts (N=15; 6 design, 9 domain) who self-rated their expertise as 4 in at least one of the fields.

Collecting evaluations for the frames

Each expert evaluated all 40 frames through an online structured survey. Because we were interested in how frames might work in a distributed collaborative design setting, we focused experts' judgments on the question of how useful they think a frame would be if given to someone else to use for ideation. Therefore we had the

Table 2. Descriptive statistics for experts' frame ratings.

	mean	sd	max	min
Overall	0.29	0.71	1.73	-1.27
Domain	0.43	0.83	2.00	-1.29
Design	0.17	0.78	1.63	-1.5

Table 3. Top 3 most frequently mentioned rationale themes
(+)/(-) denotes positive/negative qualities of frames

Domain experts	Design experts
"Disagree with specific solution angels/ assumptions in the frame" (-)	"Whether the frame is in a format of brainstorming question(+) or not(-)"
"The frame provides a specific angle"(+)	"The frame provides a specific angle" (+)
"Whether the frame provides a realistic(+) angle or not(-)"	"Whether the frame is generative(+) or restrictive(-) to ideation"

experts evaluate the frames on their potential to diminish or enhance novice designers' ability to generate novel and practical solutions to the high level voters turnout challenge. Evaluations consisted of 1) a rating on a 7-point Likert-like scale from -3 (strongly diminish) to +3 (strongly enhance; 0 = no effect), and 2) an optional question about the reason for the rating. At the end of the survey, we also asked experts to explain their general criteria for rating the frames. Table 1 shows some example frames with overall ratings (not splitting by domain and design experts).

Results

EXPERTS OPINIONS OF GOOD AND BAD FRAMES CAN BE RELIABLY QUANTIFIED

Overall, there was substantial agreement among experts' ratings, Intraclass Correlation (ICC) (2,15) =0.71. Agreement within each expert type was lower but still substantial: ICC(2,9)=0.55 for the domain experts, and ICC(2,6)=0.57 for the design experts. There were restricted ranges([min, max]) of ratings given the rating scale from -3 to 3. (see Table 2). We will return to this point in the discussion.

EXPERTS HAVE COHERENT YET COMPLEMENTARY RATIONALES ABOUT WHAT MAKES FOR EFFECTIVE FRAMES

To further explore experts' opinions about the frames, we conducted an exploratory content analysis of the rationales (reasons and general criteria) they provided for their ratings. Two members of the research team conducted bottom-up open thematic coding to discover and define common themes among the rationales. A total of 15 rationale themes were discovered.

Tables 3 shows the top 3 most frequently mentioned rationale themes for domain and design experts, respectively. Overall, both sets of experts agree that frames should be specific; but they also have complementary emphases, with domain experts emphasizing feasibility and specific solution angles/assumptions, and design experts emphasizing "generativeness" and whether the frame was in the format of a brainstorming question.

Summary

In summary, we find that experts' opinions about whether frames are good/bad can be reliably quantified, as evidenced by their quantitative agreement scores. There also seemed to be common and complementary rationales for what makes for good/bad frames.

Study 2: (How) Do Experts' Quantified Opinions about Frames Predict their Effects on Subsequent Ideation?

We now examine whether experts' quantified opinions are predictive of how those frames will help/harm the idea generation of a different set of brainstormers.

Method

PARTICIPANTS

245 workers from Amazon Mechanical Turk participated in this study. They were paid a wage of $8/hr.

EXPERIMENT DESIGN AND PROCEDURE

Participants generated ideas for our youth voting problem, randomly assigned to one of two conditions: "with frames" (N= 209), and "no frames" (N=36; to assess whether better frames are only relatively better than bad frames, or also better in absolute terms). In the "with frames" condition, participants were randomly assigned one frame from the 40 frames in our study. We manipulated the assignment of frames so that there

Table 4. Examples of idea (paraphrased for space) evaluation (mean of all raters) Quality(Q): -3 to 3; Feasibility(F): 1 to 7; Novelty(N): 1 to 7

Idea 1: "Removing the requirement to actually visit a polling station to vote by creating an app that uses SSN for identification" ***Evaluation:*** <u>Q = 3; F = 3.67; N=4</u> (Idea 1 was rated as very useful on solving the problem, slightly worse than normal in feasibility and normal in novelty)
Idea 2: "Explaining the importance of voting and tell them that voting is a right that every citizen should exercise." ***Evaluation:*** <u>Q = 0.67; F = 7; N = 2</u> (Idea 2 was rated as better than having no effect on solving the problem, extremely feasible but not novel)
Idea 3: "A reconstitution of our voting system" ***Evaluation:*** <u>Q = -0.33; F = 1.67; N = 5</u> (Idea 3 was rated as potentially worsening the problem, not feasible but pretty [illegible]

were at least 4 participants generating ideas on the same frame. In the no frames condition, participants generated ideas for the original problem statement.

Participants first brainstormed freely for 5 minutes to come up with as many ideas as they could for the problem (or specific framing, depending on condition), and then spent 2 minutes elaborating on a single idea they thought was most promising (we call it "elaborated idea" in following sections).

MEASURES

Idea outcomes: Novelty, feasibility, and quality

We want to know whether frames led to "good" elaborated ideas, operationalized by the standard creativity subdimensions [4] of *novelty* (how different is this idea from existing approaches?), *feasibility* (can the idea be implemented?), and *quality* (if implemented, would it actually help?).

Novelty, feasibility, and quality were evaluated by 3 millennial US citizens (age 19-21) who have had at least some opportunity to vote (whether or not they took that opportunity). As the "target users" for the problem, these raters have appropriate content knowledge to judge novelty, feasibility, and quality.

Each rater evaluated all elaborated ideas for each of the dimensions on a 7-point Likert like scale: 1 (everyone is doing it now) to 7 (really novel) for novelty, 1 (absolutely impossible) to 7 (really easy to be implemented) for feasibility, and -3 (will largely decrease youth voter turnout) to +3 (will really make a huge positive difference) for quality. Inter-rater reliability was substantial for novelty, ICC(2,3)=0.55, and high for feasibility (ICC(2,3)=0.70) and quality (ICC(2,3)=0.80). Each elaborated idea's final score on each dimension was the mean rating across the 3 raters. Table 4 shows some example elaborated ideas from our data (with associated idea outcome scores).

Process measures: Quantity and focus of ideas

We were also interested whether frames led to helpful ideation processes (measured by the *quantity* of ideas, and how *focused* the ideas generated during the 5 minutes free brainstorming step before elaboration on a single idea). Quantity was simply the number of ideas generated during the brainstorming step. Focus was measured by the mean pairwise similarity between ideas in the brainstorming step (as measured by Latent Semantic Analysis, a computational linguistic method for evaluating similarity between textual documents [5]), ranging from 0 (no similarity, on average) to 1 (almost identical ideas).

Combined idea outcomes with process measures, we considered novelty, feasibility, quality, diversity and quantity of resulting ideas as the 5 measurements of the performance of frames.

ANALYSIS DESIGN

Splitting frames into "best frames" and "worst frames"

We analyzed the performance of frames at the participant level, answering the question: "when frames of a certain expert rating level are given to a user, does it tend to yield better/worse ideas?" To maximize our ability to detect differences between good/bad frames, we compared the performance of frames in the bottom ("worst frames") and top quantiles ("best frames") of ratings for domain and design experts ratings separately. There were 11 worst (with N=60 brainstormers) and 11 best (with N=53 brainstormers)

In figure 2-4, the horizontal lines and blue bars stands for mean scores in “no frame” conditions and their 95% confidence intervals. In figure 2-5, all the error bars are 95% confidence intervals.

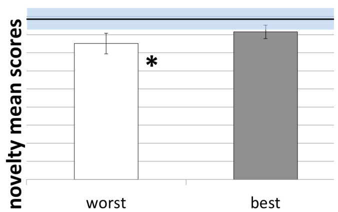


Figure 2: domain experts’ frame ratings predicted low performing frames in resulting idea novelty

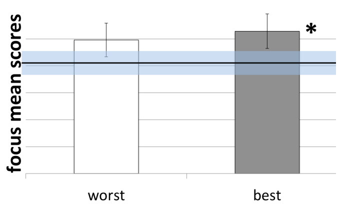


Figure 3: domain experts’ frame ratings predicted high performing frames in resulting idea focus

frames for domain ratings, and 14 worst (N=71 brainstormers) and 10 best frames (N=47 brainstormers) for design ratings.

Comparing resulting idea scores of different frame groups

We performed Welch Two Sample t-tests for the following combinations within both design and domain-based groupings: 1) best vs. worst frames, 2) best vs. no frames, 3) worst vs. no frames, and 4) with any frames vs. no frames. For example, we compared the novelty scores of participants who received “domain-best” frames (i.e., with top quantile frame ratings from domain experts) to the novelty scores of participants who received “domain-worst” frames (i.e., with bottom quantile frame ratings from domain experts). We did the same for feasibility, quality, quantity, and focus.

Results

DOMAIN EXPERTS CAN WEED OUT NON-NOVEL FRAMES

Participants who received frames that were poorly rated by domain experts generated significantly less novel ideas than participants in the “no frame” condition (see Figure 2), t(87)=-2.2, $p < 0.05$, suggesting that domain experts can predict which frames will do poorly on novelty. There were no significant results for design experts’ ratings on this measure.

DOMAIN EXPERTS CAN PREDICT FRAMES THAT HELP IDEATORS GENERATE MORE FOCUSED IDEAS

Participants who received frames that were highly rated by domain experts generated significantly more focused ideas than participants in the “no frame” condition (see Figure 3), t(83)=2.5, $p < 0.05$, suggesting that domain experts can select frames that help ideators focus their thinking. There were no significant results for design experts’ ratings on this measure.

DESIGN EXPERTS CAN SELECT GENERATIVE FRAMES

Participants who received frames that were highly rated by design experts generated significantly more ideas than participants who received frames that were poorly rated by design experts (see Figure 4), t(97)=2.4, $p < 0.05$, suggesting that design experts can select frames that generate many ideas. There were no significant results for domain experts’ ratings on this measure.

USING ANY FRAME IMPROVES IDEA FEASIBILITY

The feasibility scores of the participants who received a frame was significantly higher than the participants under “no frame” condition (see Figure 5), t(49)=2.8, $p < 0.01$, suggesting that framing brainstorming questions can help reduce infeasible ideas.

NO SIGNIFICANT EFFECTS FOR IDEA QUALITY

There were also no significant results for idea quality. For both design and domain experts rated frames, we had $p > 0.05$ in the scores of quality with any combinations of the frame conditions. This suggested neither domain experts nor design experts predicted idea quality through frame ratings

Discussion

In this paper, we explored how distributed design teams can effectively frame problems, focusing on whether and how expert knowledge can help teams select effective frames. Overall, we contribute three sets of findings for these questions.

First, we find that experts’ opinions about what makes for effective frames can be reliably quantified, as evidenced by quantitative agreement metrics. Analysis of experts’ rationales for their quantitative ratings also showed that domain and design experts had

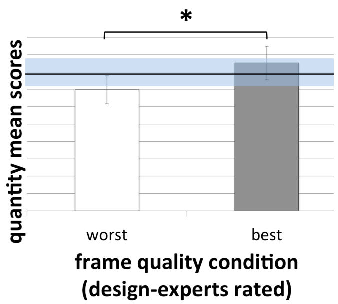


Figure 4: design experts' frame ratings predicted high/low performing frames in resulting idea quantity

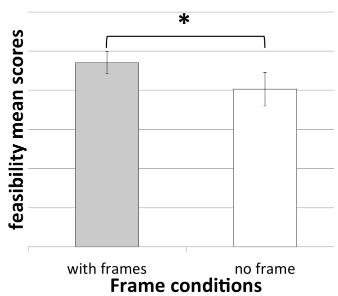


Figure 5: frames led to resulting ideas that were more feasible than ideas generated without frames

complementary perspectives to offer on what makes for good frames. All the experts agreed that effective frames are specific; domain experts added that an effective frame should also be realistic and mesh with their ideas about what makes for good solutions, while design experts thought that an effective frame should in addition to be generative and remain in a format of brainstorming question.

Secondly (and importantly), we find that experts' quantified opinions can predict in advance which frames will do better/worse when given to others to use for idea generation, although different types of experts are good at predicting different effects of frames (domain experts can help weed out non-novel frames and select frames that focus ideation; design experts can help select frames that can lead to many ideas).

Finally, we confirm prior work on the importance of framing in design: we show that framing can lead to more feasible ideas compared to simply generating ideas for a high-level problem as given.

In summary, our results suggest that expert knowledge can indeed improve the selection process for problem framing. Thus, distributed design teams might consider incorporating appropriate expert knowledge during the early stages of the framing process, depending on what dimensions of ideation are important to them (e.g., focus on domain experts if novelty and focus are important, and design experts if quantity is important).

Limitations and Future Work

Collecting frames with a wider range of quality From the frame evaluations, we noticed that we had a restricted range of frame quality: no frames were rated as strongly diminishing ideation or as perfect. To get a boarder set of frames that may potentially lead to more differentiable outcomes of ideas, in the future we want to collect more "extreme frames": e.g. frames by domain experts and frames by novices who are not familiar with the high level problem at all.

Frame ratings with metrics of rationales

Experts gave a single score for each frame in this study, but from Study 1 we learned that they had different opinions on what aspects of the frames mattered for frame quality. In the future we will expand the single score frame rating to a metric rating with the same metrics as the themes of rationales. For example, we will have the experts give scores to a frame on its specificity, feasibility and potential to generate more ideas respectively. This way we can see how ratings in different metrics predict the idea generation.

Experts to rate the ideas

In this study, the frames were rated by experts but the ideas were rated by target users (young people). Although the idea raters were the target users of the solutions, there might be some disconnects of opinions between the frame raters and the idea raters. For example, a novel solution to young people may seem not novel at all to domain experts because they have more experience and knowledge to the problem. Therefore, in the future we will recruit experts to rate the ideas as well.

References

1. Dorst, K. 2011. The core of "design thinking" and its application. Design Studies, 32, 6, 521–532. http://doi.org/10.1016/j.destud.2011.07.006

2. Dorst, K., & Cross, N. 2001. Creativity in the design process: co-evolution of problem–solution. Design Studies, 22, 5, 425–437.
3. Dubberly, H., & Evenson, S. 2008. On Modeling: The Analysis-synthesis Bridge Model. Interactions, 15, 2, 57–61. http://doi.org/10.1145/1340961.1340976
4. Runco, M. A., & Jaeger, G. J. 2012. The Standard Definition of Creativity. Creativity Research Journal, 24, 1, 92–96. https://doi.org/10.1080/10400419.2012.650092
5. 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.