Large-Scale Needfinding: Methods of Increasing User-Generated Needs From Large Populations

Cory R. Schaffhausen[1]
Department of Mechanical Engineering,
University of Minnesota,
Minneapolis, MN 55455
e-mail: schaf390@umn.edu

Timothy M. Kowalewski
Department of Mechanical Engineering,
University of Minnesota,
Minneapolis, MN 55455
e-mail: timk@umn.edu

Understanding user needs and preferences is increasingly recognized as a critical component of early stage product development. The large-scale needfinding methods in this series of studies attempt to overcome shortcomings with existing methods, particularly in environments with limited user access. The three studies evaluated three specific types of stimuli to help users describe higher quantities of needs. Users were trained on need statements and then asked to enter as many need statements and optional background stories as possible. One or more stimulus types were presented, including prompts (a type of thought exercise), shared needs, and shared context images. Topics used were general household areas including cooking, cleaning, and trip planning. The results show that users can articulate a large number of needs unaided, and users consistently increased need quantity after viewing a stimulus. A final study collected 1735 needs statements and 1246 stories from 402 individuals in 24 hr. Shared needs and images significantly increased need quantity over other types. User experience (and not expertise) was a significant factor for increasing quantity, but may not warrant exclusive use of high-experience users in practice. [DOI: 10.1115/1.4030161]

Keywords: user, needs, preferences, problems, quantity, quality, needfinding, product, design, expert

1 Introduction

The success of a new product is often determined by the degree it satisfies customer needs and preferences. However, obtaining this mix of technical, personal, and emotional content from diverse user groups is challenging, resource intensive, and often results in an incomplete understanding of a group of users. This can be particularly evident in areas with complex, often conflicting stakeholder needs ranging from energy efficiency to water and food security. Specific areas such as health care and medical devices face additional barriers such as limited access to users and user environments. A series of three studies evaluated several methodological elements of collecting needs from large groups, with a motivation to consider proven elements of effective ideation, such as a focus on quantity, and test whether similar methods can be applied to needfinding. An improved needfinding method can potentially reduce costs and time to acquire user input, include a broader range of user perspectives during development, and more effectively direct product development resources to areas of unmet user needs.

1.1 Needfinding and User Requirements. Research into product design and development strategies has identified numerous techniques to identify and understand user needs, such as needfinding [1] (also labeled needs finding [2] or problem finding [3,4]), user research, market research, or ethnographic research. These often are also grouped within the umbrella of user-centered design [5] or as voice of the customer [6] and are described in disciplines ranging from product development [7] to business management [8,9].

Needfinding is an element of user-centered design used to inform early development phases [7]. The objective of studying the user is to understand what unmet needs exist and how these needs can inform the requirements of new products [1]. One dominant theme in needfinding is to go straight to the group of users itself. Often product failures can be traced to a faulty over-reliance on input from company managers or designers rather than information directly validated with users [8]. Validating these assumptions often requires prolonged engagement to develop a deep understanding of the users' actual behavior, because actions can differ from what is said. This engagement also facilitates empathy for users, and empathy is critical for recognizing the needs and differing perspectives of users [8,10–12]. The objective of needfinding is to consider only needs, agnostic of solutions, and to be mindful that statements seemingly reflecting needs can include embedded solutions [1,2].

The engagement with users typically occurs with observations and in-depth interviews, which might be described as methods for ethnographic research [13] or qualitative research [14]. Observation studies focus on what is done rather than what is said. They do not require a user's conscious awareness of a need in order to capture it [1,13,14]. Qualitative interviewing is a form of interviewing that relies on open-ended questions to allow for depth and completeness in answers where there appears to be the opportunity to uncover important insights. The questions are carefully directed by the interviewer to allow the subject to give a thorough report from the subject's point of view [15,16].

In some cases, the process for identifying needs is intentionally divergent, to identify a large pool of potential needs. Griffin and Hauser [17] used consumer products data to develop a function for the increasing proportion of total needs with increasing user group size. They suggest a range of 20–30 1-hr interviews with different individuals with data reviewed by multiple (up to 7) analysts in order to identify approximately 90–95% of possible needs [17]. While this study does not include user observations and is a single study, it remains a commonly cited baseline. A filtering, or

[1]Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received September 11, 2014; final manuscript received February 13, 2015; published online May 19, 2015. Assoc. Editor: Carolyn Seepersad.

 Copyright © 2015 by ASME

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

convergent, process follows and may be largely data-driven, for example, based on market size, development costs, etc., or may be similarly influenced by personal factors such as individual interests and motivations [2]. Bayus [18] provided a thorough literature review on how this phase feeds into subsequent phases such as ideation for solutions.

Using an understanding of users to develop successful products remains challenging, in particular when developing radically new products [19,20]. Within consumer products, purchasing decisions and shopping experiences may be significantly affected by hedonic value [21,22], and products can include complex emotional content [4] as well as symbolic meaning (e.g., evoking luxury or personal aspirations) [23]. In contrast, specialized areas such as medical device purchasing are increasingly institutionalized and data-driven, suggesting a greater importance of needs [2].

1.2 Lessons From Ideation for Needfinding. Needfinding shares several core challenges with ideation. Decades of research have largely addressed these in the domain of ideation, and the further study of remaining challenges in needfinding may benefit from lessons of ideation research. Ideation has been studied in the context of product design [24] and problem solving in general [25]. It is a divergent process used to generate a large pool of ideas, typically focused on solving a specific problem. In this sense, needfinding could be considered similar to ideation, although to identify many needs, rather than solutions.

Brainstorming is arguably the most prominent technique used today for ideation and dates to work by Osborn [26]. His work described the brainstorming process as an interactive group activity with a goal of generating a large number of ideas in a short amount of time. Some key procedures are to focus on quantity, encourage building off of the ideas of others, and to withhold criticism of other ideas or members. He also hypothesized that simply generating more ideas will lead to more good ideas [26].

In the years since, Osborn's hypothesis on quantity to achieve quality has been extensively studied. Brainstorming research shows clear evidence that the specific techniques of ideation or brainstorming have a significant effect on the result. Subsequent research has evaluated a variety of brainstorming methods to improve outcomes including several software interfaces to mediate group interactions [27–29], and the results support the benefits of a software interface for achieving high quality and quantity in ideation [30–32].

Evidence also suggests that the content of the instructions used to begin the session can impact the results. Paulus [33] replicated previous results showing that instructions to "generate as many ideas as possible" improve quantity and quality results relative to control groups and groups given instructions to focus on "high quality ideas" [33, p. 41].

A trend has emerged from the general body of research supporting Osborn's hypothesis, namely, that there is a correlation between quantity and quality of ideas during brainstorming. There are a large number of constructs suggested in literature for evaluating the quality of ideas [24,25], yet the correlation between quantity and quality has been affirmed both for cumulative group quantity [33–35] and also individuals within a group [24].

As described in Sec. 1.1, needfinding methods suggest going straight to the source, and this often means targeting expert users [8]. On the other hand, it is relatively common to brainstorm on a solution to a problem without being an expert on the problem. Evidence from crowdsourcing ideation platforms such as InnoCentive® suggests that outsiders to a specialized field can make connections across disciplines and suggest innovative solutions that were not evident to experts [36].

1.3 Users Articulating Needs. A common perception in needfinding literature is that users typically are not able to directly articulate their own needs when asked. This is often attributed to the formation of habits or dogma, as users adapt to shortcomings and no longer recognize them or are not easily able to see beyond the existing set of solutions [2,8]. Researchers studying ideation have struggled with similar effects, often described as design fixation [37]. While cognitive mechanisms may not be identical, it is worth noting that the presence of fixation in ideation is seen as an obstacle to be overcome, not a fundamental flaw of the method [38,39]. The scrutiny in research shown in ideation has not been directed toward testing methods to overcome this fixation-type effect in needfinding.

Many methods to understand user needs focus on improving a designer's ability to identify someone else's needs [18,40]. Few methods have been rigorously tested to help a user better articulate his own needs, although successful methods such as empathy tools have been reported in a trial study [41]. The use of crowds appears extensively for ideation ("open innovation") and later phases of product development [36,42]. Crowd data have been used for needfinding and user preference modeling; however, this work employs data mining of existing content such as blog posts and comments rather than direct solicitation of needs [43,44]. A gap remains for needfinding applied both to crowds and to directly solicit needs from users. In spite of a lack of prior research targeting crowd-submitted needs, Faste [45] explicitly stated that advances in online knowledge management "could be applied to crowdsourced needfinding research" [45, p. 5], and further observes "Perhaps one of the most important ways in which open-innovation can therefore be made to thrive is by enabling individuals to report their own needs" [45, p. 4].

1.4 Rationale for Large Quantities of Needs. As described in Sec. 1.2, a correlation does exist between high quantity and high quality in ideation, and this process shares a similarity with needfinding in that the desired outcome is a pool of potential candidates to pursue further. While this correlation in needfinding was not tested in these studies, a method to collect large quantities of needs would be a necessary first step. In addition, given the challenges identified in Sec. 1.1, uncovering a unique need can be a rare event and a higher quantity of attempts would have a higher likelihood of a rare event occurring. Ultimately, what matters is the count of high-quality needs, not necessarily the proportion. A final output of 10 high-quality needs and 500 poor-quality needs would be superior to 2 high-quality needs and 50 poor-quality.

1.5 Rationale for Our Stimuli Design. Three types of stimuli were tested in these studies. The shared needs stimulus was intended as analogous to ideation and brainstorming sessions where participants are primed with the ideas generated by others. Dugosh and Paulus [35] have evaluated this method for ideation, which assumes that exposure to ideas from others will stimulate new ideas. The positive effects of shared ideas in ideation have also been reported for increasing the number of ideas categories [46], increasing idea generation in electronic brainstorming sessions [47], and increasing combinations based on shared ideas [48]. Verbatim shared idea content is not required, in fact, subtle encouraging cues are also sufficient for increasing idea generation [49].

The contextual image stimulus was intended to help provide context to the activity, as most participants might be at a computer far removed from an environment related to the topic. Availability of context is previously described as a key rationale for observational study. Retaining contextual information may potentially trigger useful insights [1,50]. Visual examples have been used previously for priming and mitigating fixation in ideation, and the effects have been positive as well as negative [37,39,51]; however, this study presented images for a more general purpose. The images were assembled to be more than a set of visual examples of problems. They represented broader, general context for the topic area.

The prompt stimulus was intended as a substitute for probing questions present in qualitative interviews. Some allowed

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

participants to focus on particular elements of products, for example, a specific faulty or broken product. Others allowed participants to focus on particular events, such as a recent emotional experience. The objective was to facilitate self-reflection or thoughts of empathy.

1.6 Application Example: Heath Care and Medical Devices. Large-scale needfinding can benefit a variety of application areas. As an example, we herein discuss medical device development. In addition to complex stakeholder groups, this area faces additional challenges such as restricted access to user groups. Traditional needfinding methods such as in-depth or immersive observations face greater barriers to accessibility compared to consumer devices [52]. In addition, the highly regulated nature of medical devices increases the cost and time of development projects [53], and therefore, increases the risk to companies if the product is unsuccessful.

Researchers commonly noted that known formalized methods are not consistently used within the constraints of actual industry development projects [54–57]. Money et al. documented a series of in-depth interviews with industry management and described deficiencies such as primarily consulting physicians or surgeons in spite of identifying the end user as a patient or nurse. Even when device users were professionals, interviewees expressed a preference for informal methods of capturing user requirement input from "a small number of esteemed medical experts" [54, p. 11]. A different survey found that while "informal expert review" was among the top five methods used, it was ranked as one of the least effective [58]. A third survey, extending beyond health care, supported these findings. Here, methods such as ethnography were ranked as the most effective, but saw limited use in practice [6]. In addition, previous studies have highlighted the improved outcomes resulting from understanding stakeholder needs from diverse groups [59].

A wide range of barriers has been observed, which prevent both the use of formal methods to assess user requirements and the implementation of findings [52,54,60]. Barriers include cultural beliefs of management [61], limited resources for time intensive methods, lack of expertise of methods [52,54], lack of accessibility of users, [52], and uncertain value of qualitative results [54,60].

New methods could potentially overcome several of these barriers through remote interactions and shorter time commitments for individual users. Given limitations for access to health care professionals, testing a sequence of new methods initially with professional groups was not feasible. Validation work engaged alternate groups of users to prepare for future case study work in target application areas.

2 Methods Overview

Three sequential studies were completed to evaluate different aspects of the needfinding method. Study methods differed by necessity to address different objectives as shown in Table 1. Each study included an online user interface to collect open-ended need statements. This interface was combined with a method to display stimulus information to potentially help increase the quantity of needs a user could articulate.

All three studies asked users to submit single-sentence statements describing problems or unmet needs relating to a single topic. After entering a need statement, a participant could enter a more elaborate story to describe relevant background information. Each participant was randomly assigned to one of three topic groups: preparing food and cooking, doing housecleaning and household chores, and planning a trip. The topics were selected to be familiar to a majority of individuals recruited from online communities and provided variation in the types and nature of products and services that might be discussed.

Table 1 Summary of study objectives

Study 1	Test matrix of prompts to increase need quantity
Study 2	Compare control, prompts, shared needs, and shared images to increase need quantity
Study 3	Test unstructured availability of three help types to evaluate a case study scenario

All participants were recruited from the Amazon Mechanical Turk (AMT).[2] AMT is a site allowing a community of task requesters (analogous to employers) to recruit individuals from a community of online workers. The tasks are divided into discrete deliverables called Human Intelligence Tasks (HITs), and workers are paid nominal amounts for each deliverable. Pay is generally proportional to task duration and falls within a broad range of approximately $10 per minute. AMT is increasingly used as a source for research participants, and the user population has been previously characterized [62]. Participants recruited from AMT were directed to a custom survey interface developed using ZOHO CREATOR.[3] ZOHO CREATOR is a cloud-based custom database platform with integrated logic scripting and graphic user interface development tools.

Participants had to meet basic requirements in order to be eligible. These included approval rates of 95% or higher for completed work, a history of at least 100 completed HITs, and a United States IP address location. Each study allowed repeat workers who had previously completed an earlier study. In this case, the worker would automatically be assigned to a different topic area than any previously seen.

AMT workers would accept the HIT and would see that the objective was to describe problems with common products and services. Instructions were framed in a variety of ways that might be clear to a wide range of people. For the topic of cooking, instructions included: "We want to know what would make preparing food and cooking a better experience. Examples can be very broad, for example: more convenient, less effort, safer, easier to understand, cheaper, more consistent, or faster." (Examples differed for each topic) "You will type in descriptions of problems or unmet needs you face preparing food and cooking." "You want to describe these so someone could make improvements or offer solutions in the future. Try to think of as many as you can."

After reviewing consent information, participants completed a training exercise. Training began with brief instructions stating that inventions should not be included and to describe the problem in a complete sentence. Participants then took a quiz including five example statements and were required to identify which were not consistent with the instructions. The examples and quiz related to a new topic (reading books) to avoid providing example needs relevant to assigned topics. The training was paid as a fixed amount of $0.65 for both pass and fail outcomes. Participants who failed were not able to continue. Following the quiz, participants answered optional demographics questions including gender, age, and self-reported levels of expertise and experience (hours per week). When training was complete, participants began entering needs and stories. The final instructions, again for the topic of cooking, were "Don't worry about whether the benefit is worth the cost. We simply want lots of suggestions." Each entry was paid as an individual bonus. Bonus amounts varied for different studies, as shown in Table 4.

Studies 1 and 2 were designed to differentiate between needs readily available to the user and those that may have arisen as a result of viewing some type of stimulus information. These studies presented users with two options in the interface: "Enter Another" and "I'm Stuck" buttons. The I'm Stuck button was described as the option if the participant was not sure what to say. The intention was to treat initially submitted needs as available to the user simply when asked, and needs submitted after pressing

[2]https://www.mturk.com/
[3]https://creator.zoho.com/

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

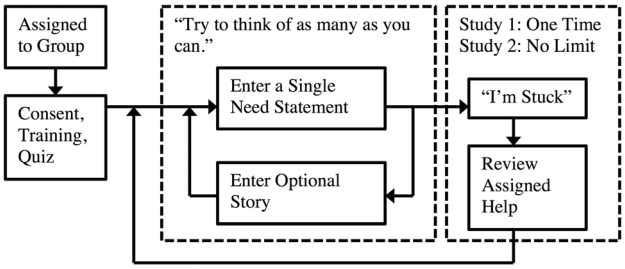

Fig. 1 Summary schematic of study 1 and study 2

I'm Stuck as potentially generated through viewing the stimulus. The stimulus was described to users as "help," and the available help differed for each study. For this discussion, "stimulus" and help can be considered interchangeable. Table 1 describes the types of help available for each study. After viewing a stimulus, the user returned to the interface to enter any new needs and stories. A general process schematic is shown in Fig. 1.

2.1 Study 1 Methods. Study 1 tested the effectiveness of a single type of stimulus, a paragraph-length narrative prompt, described in more detail in Sec. 2.5. The effectiveness of a type of stimulus was measured using a count of needs submitted after selecting I'm Stuck and reviewing the prompt. Participants in study 1 who clicked I'm Stuck twice were shown a message that only a single help was available and then the study ended.

Study 1 screened a total of 30 prompts (including a control). The study employed a sample size of 15 users per prompt. Study 1 data were analyzed to identify if any prompt or trait of prompts resulted in a lower mean of needs submitted after viewing. These prompts could be omitted in future studies. Prompt traits were analyzed by grouping prompts along rows or columns of the complete matrix described in Sec. 2.5. A likelihood-ratio test was used to determine the best fit model comparing Poisson and negative binomial models. A regression analysis was used for the best fit model to test differences of groups (models tested differences of log(means)). A multiple comparison test (multcomp R package using "Tukey" parameter) was used on the generalized linear model to test pairwise combinations of prompt matrix rows and columns [63].

The needs count data from study 1 were used to calculate a sample size for groups in study 2. For an initial approximation, the distribution was assumed to be a Poisson distribution despite some evidence of over-dispersion. The approximate sample size would be dependent on the assumed group means, with a desired delta of 1 need per person for differing groups. Table 2 shows a range of assumed means with this delta value and the resulting range of sample sizes.

The assumed rate of failed training was 50% based on previous studies. However, given the uncertainty in the pass/fail rate and the potential for exiting the study prematurely, a conservative target sample size for study 2 was 100 per group. This exceeds the calculated sample sizes shown in Table 2. In order to achieve this group size for passing and complete responses, study 2 recruited 150 participants per treatment group.

2.2 Study 2 Methods. In this study, three types of help were available and are listed in Table 1. Details for each type are given in Secs. 2.4–2.7. The first time a user selected I'm Stuck, the randomized help was selected from the three types and control group. In study 2, selecting I'm Stuck a second time allowed participants to begin to select any additional help at will. Study 2 analysis used the same metric for effectiveness of a stimulus, specifically, the number of needs submitted after viewing the stimulus. Only needs entered after viewing the first help but before viewing any subsequent help were included in this metric.

Table 2 Study 2 target sample sizes for selected group means

Θ_1, group 1 mean (needs)	1	2	3	4
Θ_2, group 2 mean (needs)	2	3	4	5
$n^{a} = \left(4/\left(\sqrt{\Theta_1} - \sqrt{\Theta_2}\right)^2\right)$, sample size	24	40	56	72

[a]Poisson distribution.

Study 2 data were analyzed to test for a significant effect of stimulus type (prompts, shared needs, and images). Statistical tools were identical to study 1, as described in Sec. 2.1.

2.3 Study 3 Methods. Study 3 differed from the process shown in Fig. 1 by providing options to view any of the available help from the beginning, omitting the I'm Stuck button. This study randomly assigned participants into topic groups, but did not assign participants into any test groups. Histograms, empirical cumulative distributions, and descriptive statistics were used to make basic observations such as the number of times participants would choose to view additional help and the resulting number of needs submitted. Study 3 used an interface relevant to a case study application where options to quit or receive ongoing help were readily available. Figure 2 is a sample image from study 3. This screen was consistently used for entering needs. The requested help information was displayed in an adjacent box (not shown) to allow simultaneous display. The target sample size for each topic was approximately 125 per topic group, resulting in a similar total size compared to study 2.

2.4 Control Stimulus. Studies consisting of a control group used nominal additional bonuses, for example, a "double bonus," to encourage continued participation and potentially limit the rate of quitting prior to reviewing the stimulus information. A control group would be offered only this additional bonus, and each treatment group would be offered the same additional bonus and also a display of stimulus information. This additional bonus was not considered a treatment, as it was consistent for all groups and effects of incentive were not tested.

2.5 Stimulus 1: Narrative Prompts. The first type of stimulus was a prompt to ask users to think about a particular task from different perspectives. This focus may help identify a particular type of need. The prompts were arranged in a matrix to organize these perspectives based on similar traits. For example, one axis of the matrix related to differing contents, such as a focus on a particular emotion (e.g., frustration) or type of communication (e.g., instruction manuals). The other axis of the matrix related to different subjects, such as a first-person view or a third-person view. Traits were derived from design empathy literature [8,10,11] and interviewing methodology [15,16] and combined with new variations. Each cell of the matrix contained one or more combinations of these traits. Figure 3 shows an outline view of the matrix rows and columns.

This type of stimulus included a total of 29 prompts combining a variety of traits described above. The same group of prompts was used for all studies. Table 3 shows an example of a complete prompt. The complete details of traits, prompt matrix cells, and prompt content are available upon request.

2.6 Stimulus 2: Shared Needs and Stories. The second type of stimulus allowed a user to read the entries submitted by previous users. Study 1 was used to collect this pilot data. Needs submitted in study 1 were reviewed, and incomplete sentences and inventions were omitted. Only needs with an accompanying user-generated story were shared in studies 2 and 3. The complete group of needs was randomly ordered and grouped into batches of ten needs and ten corresponding stories. The total shared needs content included approximately 30 batches available for each

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

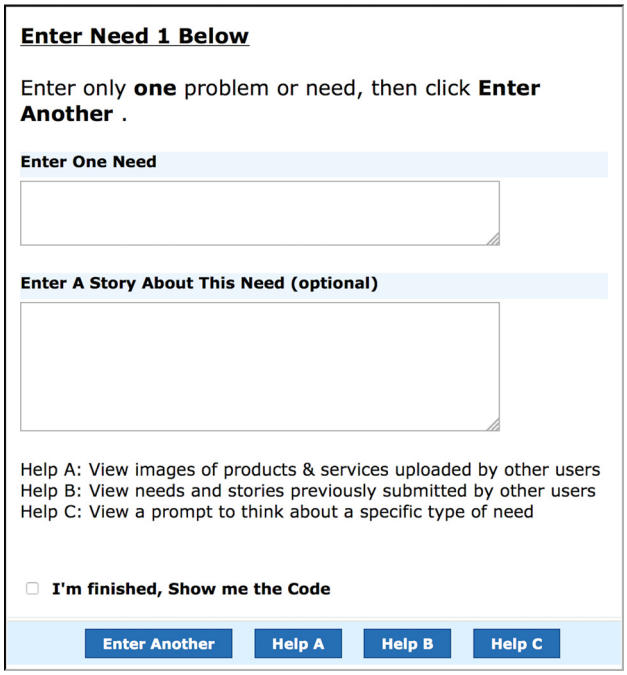

Fig. 2 Study 3 user interface for entering needs

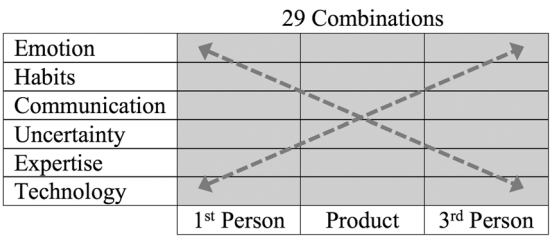

Fig. 3 Summary outline of prompt matrix

topic. Table 3 shows an example of a need/story pair selected as one with a particularly vivid description.

2.7 Stimulus 3: Shared Images. The final stimulus was a display of a series of content-specific images submitted by previous users. An independent pilot was used to collect these images. The pilot was repeated twice, each time assigning a participant to one of the same topic groups used for studies 1–3. In the first iteration, the pilot participant was asked to upload an image of a product or service used for or relevant to the topic. The second iteration asked the pilot participant to upload an image relating to something the person disliked about the topic. Images submitted in these pilots were reviewed and irrelevant images or images with faces or identifying information were omitted. The complete group of images was randomly ordered and grouped into batches of ten. The total shared images content included ten batches for each topic. Table 3 shows an example of a portion of one shared image included in a batch for planning a trip.

2.8 Stimulus Instructions. Users assigned to or requesting the narrative prompt stimulus were instructed to read the passage and see if thinking about the topic in this way resulted in any new needs. Users assigned to or requesting shared needs or images were instructed to review the complete batch as inspiration for a type of brainstorming activity and to think of new needs related to the shared information or anything new that comes to mind. Each participant was shown a random batch corresponding to the assigned topic. The participant never viewed repeated content (the same prompt or image), specifically when repeat help was available.

3 Results

Summary data for each study are presented in Tables 4–6. Table 4 includes data relating to the AMT system and worker payments. In total, approximately 1730 workers were paid for completing a study. Table 5 includes data relating to the ZOHO survey database. In total, the three study surveys were accessed approximately 2300 times. The disparity in participant counts between the AMT data and the ZOHO data is due to multiple exit points during the survey, which were prior to completing training and getting an authorization code to be paid by AMT. For example, many

Table 3 Examples of each stimulus type

Example narrative prompt for cooking	Example shared need and story for cleaning	Example image for planning a trip
"Think of a time when you tried preparing food and cooking, and the result did not end up how you had hoped or wanted. You were expecting to get a certain result, but that isn't what happened. Can you identify any reasons why you didn't get the outcome you expected? What problem could be addressed to help get the outcome you wanted?" This prompt combined a first-person view with content relating to uncertainty	Need: "I wish there was an easier way to clean the back side of the toilet that is hard to reach." Story: "The last time I cleaned the bathroom I got down on hands and knees as usual to clean the back part of the toilet. To my dismay I found that I had to hug the nasty toilet to even reach that part, and I have long arms, so I can only imagine how my wife gets back there to clean. I wish there was something to [sic] would make it easier to reach that part of the toilet without necessarily being hard on your wrists or hands or unnecessarily heavy"	

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

Table 4 Summary of AMT data

	Study 1	Study 2	Study 3
Total HITs submitted	530	600	601
Total base payments (USD, excluding Amazon fees)	$335	$390	$390
Total bonus payments (USD, excluding Amazon fees)	$228	$299	$273
Bonus for needs (USD)	$.20 for 5	$.05 ea	$.05 ea
Bonus for stories (USD)	$.10 ea	$.15 ea	$.15 ea
Study duration (days)	20	2	1

Table 5 Summary of study participants

	Study 1	Study 2	Study 3
Granted consent and began the study	775	725	810
Excluded			
Quit during training	87	96	171
Did not pass training quiz or attempted to retake	276	264	219
Passed training but quit before need entry	44	0	18
Included in analysis	368	365	402
Repeat workers (included in total)	N/A	4	57

Table 6 Summary of need and story results

	Study 1	Study 2	Study 3
Total workers submitting 1+ needs	355	347	341
Total workers submitting 0 needs	13	18	61
Total needs submitted	2441	1795	1735
Total stories submitted	1172	1332	1246
Average need length (characters)	84	73	74
Average story length (characters)	N/A	278	269
Min/median/max needs per person	0/6/68	0/4/34	0/3/28
Min/median/max minutes to enter needs and stories	N/A	2/2016/116	1/11/172
Total count of help views	N/A	483	549
Workers viewing 0 help	0	0	206

participants agreed to be in the study, but quit after reading the instructions. Table 5 provides a list of participants who were excluded from analysis due to failing training or incomplete data.

Table 6 provides an overview of needs and stories submitted with each study. In total, approximately 6000 need statements and 3750 stories were collected. Some data in Table 6 are not available for study 1. Story entry length was inaccurate because a number of participants combined multiple needs into a single entry and the accompanying story may not have described all needs. Also, study 1 did not record beginning and end times when participants were entering needs. Finally, help was offered only a single time per user in study 1.

While the future systematic assessment of unique and nonunique need submissions is described in Sec. 4.8, a preliminary review of data did not indicate malicious copying of other needs, particularly given opportunities to view shared needs. Complete need sets for studies 2 and 3 were reviewed using standard software (R) to compute total sentences and total unique sentences. In addition, each complete need set was sorted alphabetically and manually reviewed for duplicate or near-duplicate entries (e.g., missing punctuation). Potentially copied sentences were less than 1% of totals in all sets.

3.1 Study 1 Results. A negative binomial regression analysis was used for testing difference of log(means) for study 1 based on likelihood-ratio test results. The model fit is preferred over Poisson due to the presence of count data with over-dispersion.

The results of a two-sided test indicated that there were no individual prompts, rows, or columns with a significant difference lower than others. Likewise, pairwise comparisons for both rows and columns did not reflect any significant differences. The lack of isolated lower performing types of help gave no rationale to exclude any particular prompts in future studies.

Two individual prompts showed a significantly higher mean (p-values of less than 0.001 and 0.003), although the highest prompt mean corresponded to the most prolific individual (triple the count of the next highest individual). Only a single row and column mean showed a greater than marginal difference (p-value less than 0.01), and these corresponded to the row (p-value 0.0005) and column (p-value 0.004) of the most prolific individual. The median number of needs submitted after a prompt was the same for all prompts (1 need).

3.2 Study 2 Results. A negative binomial regression analysis was used for testing difference of log(means) for study 2 based on likelihood-ratio test results. The model fit is preferred over Poisson due to the presence of count data with over-dispersion.

Due to a slightly higher than anticipated rate of exclusions for failed training (see Table 5), the final group sizes included a minimum of 90 per group, rather than the target of 100.

Figure 4 compares the needs submitted after viewing a stimulus for each type tested in study 2.

The shared needs group resulted in a significantly higher mean of needs submitted compared to the control and prompt groups (p-values 0.003 and less than 0.001, respectively). The shared images group was a marginally higher mean compared with the prompt group (p-value 0.04).

3.3 Study 3 Results. Table 7 lists the number of times each type of stimulus was voluntarily requested during study 3. Voluntary selections by all users reflect the number of times each type of help was selected for the entire study population. Voluntary selections by users viewing at least one of each help type show the number of times each type of help was selected by the subset of users who had the opportunity to see all three types and would have known what type of content is shown for each.

Figure 5 shows how many needs were submitted after viewing each type of help for each request for help. The needs submitted at 0 help selections reflect those entered before viewing any help. For study 3, the maximum number of help selections was not limited, and each additional selection shown resulted in additional needs entered.

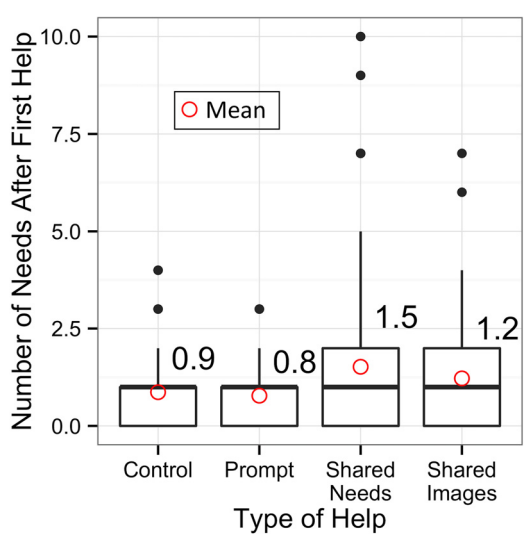

Fig. 4 Study 2 comparison of stimulus types

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

Table 7 Study 3 participant preferences selecting help

	Prompt	Shared needs	Shared images
Voluntary selections by all users	143	202	204
Voluntary selections by users viewing at least one of each	114	132	130

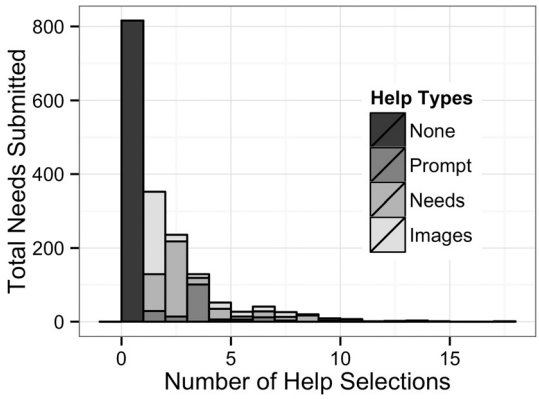

Fig. 5 Study 3 needs submitted for each help type

The cumulative distribution of needs submitted after repeated requests for help is represented in Fig. 6. The figure includes needs submitted before viewing any help, indicating just under 50% of needs were submitted before viewing help. Observe a diminishing return for continuing help requests where 90% of all needs were attained after the first three helps, and 98% were attained after eight helps.

3.4 Aggregated Observations for all Three Studies. Data were aggregated only for descriptive statistics and measuring covariate effects. The analysis of the effects of topic (e.g., cooking or cleaning) and other covariates was performed on combined data. The relative contributions of level of expertise (self-rated), experience (self-rated hours per week), topic area, and study iteration were tested with likelihood-ratio tests for negative binomial regression models. The topic area was not significant (p-value 0.42); therefore, data from all topic areas are aggregated for this analysis.

The level of expertise was not a significant variable (p-value 0.13). Figure 7 shows the total number of needs submitted per person for each study and each expertise level. The group size for differing expertise groups varied from approximately five professionals per study to approximately 200 intermediates per study.

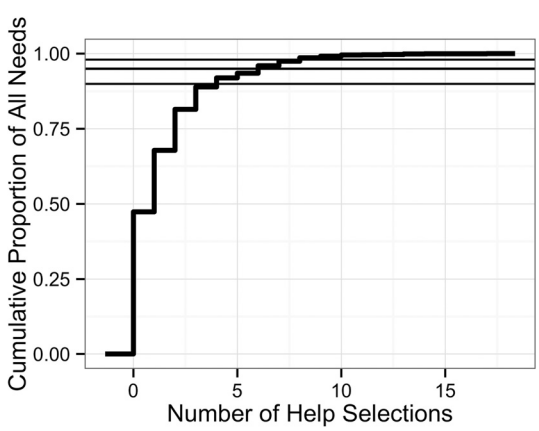

Fig. 6 Study 3 diminishing returns with increasing help (lines at 90%, 95%, and 98% are shown)

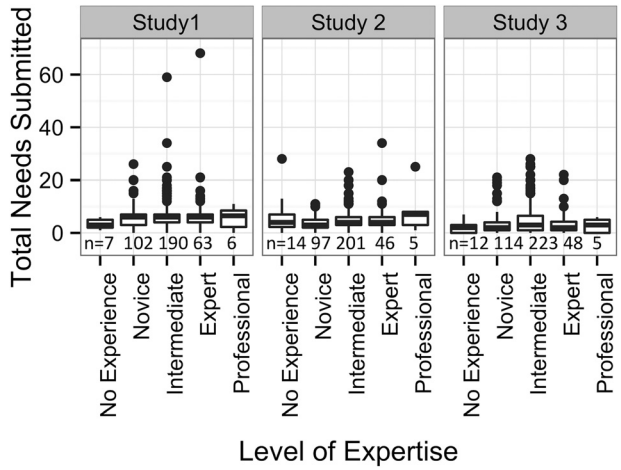

Fig. 7 Needs submitted by each expertise group (group sizes, *n*, are shown)

The level of experience (hours per week) was a significant variable (p-value 0.003). Figure 8 shows the total number of needs submitted for each study and each level of experience. The group size for differing experience groups varied from approximately 30 individuals per study with 10+ hr per week experience to approximately 220 individuals per study with up to 5 hr per week. The "Up to 5 hr" group mean was 0.9 higher than the "None" group, up to a maximum difference of 2.4 higher for "More than 10 hr" compared to None. A multiple comparison test (using only the significant model factors of study and experience) showed higher experience levels consistently resulting in higher need quantity. "5–10 hr" and more than 10 hr were significantly higher than None (p-values less than 0.001 and 0.001, respectively), and 5–10 hr and more than 10 hr were significantly higher than up to 5 hr (p-values 0.009 and 0.046, respectively).

Figure 9 shows the relative contributions of needs submitted for the complete study for each expertise level. The variation in group sizes is again evident, and descriptively, the shape of distributions for each group is very similar.

Figure 10 shows the cumulative distribution of need submission over the duration of study 2 and study 3. Study 1 was omitted as Table 4 shows that the duration of study 1 was an order of magnitude greater than the other studies. Study 2 showed a distinct change in slope at approximately 20 hr, and after this point proceeded with a rate similar to study 3. Study 3 reaches a point of

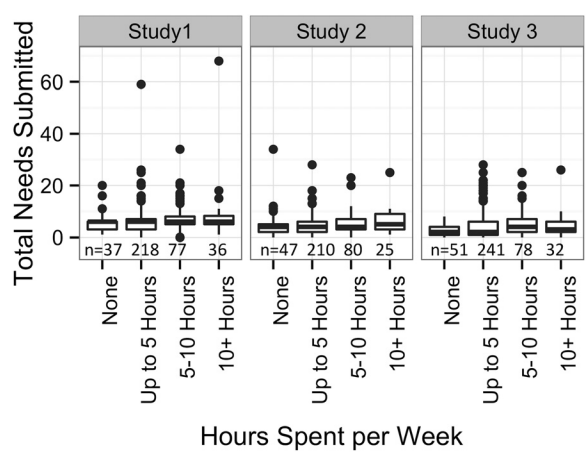

Fig. 8 Needs submitted by each experience group (group sizes, *n*, are shown)

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

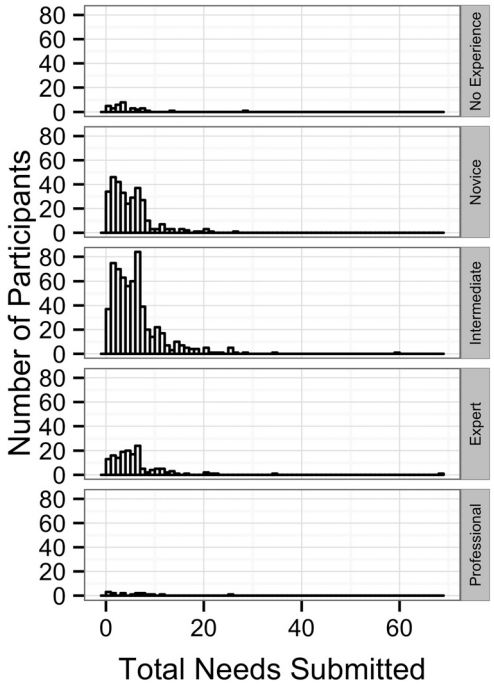

Fig. 9 Distribution of needs submitted per person across expertise groups

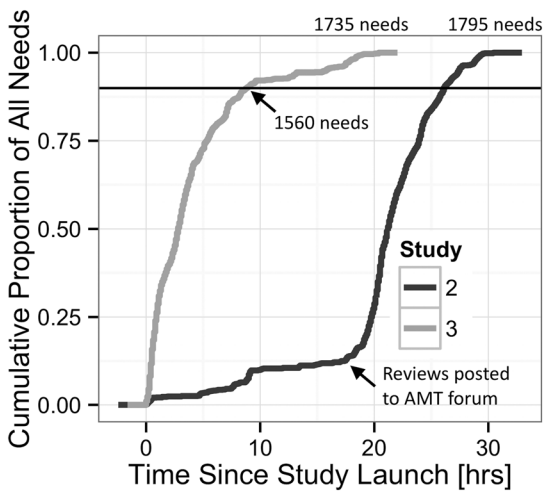

Fig. 10 Rates of need entries for studies 2 and 3 (line at 90% shown)

90% at approximately 8 hr, corresponding to an average rate of approximately 200 need statements per hour over this time interval.

4 Discussion

The goal of this study is to contribute a needfinding method allowing rapid collecting of needs from large groups. The results showed strong evidence supporting large group needfinding and spurred multiple important observations.

4.1 Fast, High Volume Need Collection Is Feasible. Figure 10 shows that in the equivalent time of 1 day of traditional ethnographic observation, an alternate crowd-based method can collect 1500 need statements and 1100 stories. This is not sufficient to suggest that this method is superior to existing ethnographic methods, only that there may be a higher rate of needs and that at a minimum, this source of input could compliment observational sources.

4.2 Collecting Needs Does Not Require Observations. These studies provide strong evidence that users will have the ability to articulate needs directly when the interaction is mediated by sufficient background and instructions, incentives, and stimuli. There is rationale to assume additional types of stimuli and incentive structures may further improve the outcome of directly soliciting needs from users. However, this does not suggest that observations should be omitted when resources and user access permits. User observations may continue to increase understanding and empathy at any phase of development and may also help identify, clarify, validate, and prioritize a set of needs.

4.3 Effects of Incentives and Stimuli. The results of these studies indicate that specific stimulus types can significantly impact the quantity of needs collected, and the incentive structure appears to influence user behavior. This was evident comparing different help types; however, study 1 results did not suggest conclusive evidence that specific traits or prompt content were a significant factor. Additional study may be necessary to identify what specific content is most effective. While needs per person for different stimulus groups may vary by a relatively small difference of means (less than 1), this method consists of aggregating needs for hundreds of individuals, and the resulting effect of the combined group could be a difference of several hundred needs. The outcome of the shared needs and shared images is a significantly higher quantity of needs; however, these types of stimuli require pilot data. Real-time sharing may reduce the dependence on pilot data for studies where all users begin at approximately the same time, but this might be less successful in an asynchronous method as used here. Although a direct financial incentive showed some positive effect as a control, some application areas will prohibit direct payment incentives, so a prompt stimulus may still be useful in absence of pilot data. This positive effect of stimuli is consistent with previously demonstrated improvements in user needs generation when providing users with empathy tools for extreme use scenarios [41].

The specific amount of bonus payments may have had an effect on user behavior. The bonus per story increased from $0.10 for study 1 to $0.15 for study 2 and study 3. The change was motivated by a goal to increase what was viewed as a valuable source of additional information. The proportion of needs submitted with stories increased from approximately 50% to a minimum of 72% after this change.

4.4 User Expertise and Experience Are Not Interchangeable. In spite of potential similarities between a user's expertise and experience, the former was not a significant variable and the latter was. One potential reason for the discrepancy could be a user's inaccuracy or bias in self-rating expertise. A sense of expertise may be influenced by multiple factors including past experience or comparisons to immediate peers. Specialized users might have expert status based on credentials rather than recent experience. In other words, an expert user may have formerly spent a significant time on the task, but no longer does. With this consideration, needfinding results might improve when prioritizing users with current experience over expert status.

However, this difference may not, in fact, warrant targeting only higher experience levels in practice. The increase from up to 5 hr (5.0 needs per person) to more than 10 hr (6.5 needs per person) is 1.5, or approximately 30%. In this case, the aggregate effect discussed for stimulus groups may not be seen here because high-experience groups generally were much smaller. The difference in mean should be considered in conjunction with other real-world factors such as overall cost as determined partly by recruiting costs. In a scenario where higher experience in users results in a 30% cost increase, this method would collect a greater

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

number of needs for a lower cost by recruiting available workers even if they are not highest in experience.

4.5 Collecting Data on AMT. Studies 1–3 were conducted as validation activities to prepare a needfinding method for use in a specific application area of clinical care delivery or training. However, the results do support the potential use for collecting large quantities of needs from the general public.

Table 4 clearly indicates that study 3 benefited from a consistent and rapid rate of need entry; however, other studies had differing results. A likely explanation for the 20 day duration of study 1 is found in the information-sharing infrastructure of crowd sourcing communities. There are a number of AMT worker forums where workers post reviews of completed HITs and rate the quality of the task and fairness of the requester regarding payments. Requesters who launch a first study have no reputation of task quality or fairness to aid in recruiting workers should they investigate a requester prior to starting. Each study was performed with a conscious effort to provide an experience worthy of positive AMT forum feedback, including setting clear expectations, a fair pay rate, and prompt payment processing. A slight increase in rate during study 1 (not shown) and a significant increase in rate for study 2 (see Fig. 10) corresponded in time to positive reviews posted to worker forums. It is likely this gradual accumulation of positive, public feedback contributed to an initial high rate of recruiting and need submission for study 3.

Also of note, Table 6 shows that study 1 actually finished with the highest count of needs regardless of the fact that help was most limited. One potential explanation again points to the influence of worker forums. Study 1 had little initial feedback posted to forums and quickly became one study in a sea of thousands of available tasks. Contrast this with studies 2 and 3 where early positive feedback gave the study high visibility among a subset of the crowd who rely partially on this input to decide which tasks to complete. It is possible that later studies will be taken by crowd workers based on factors such as a reputation for prompt payment rather than the study content. Note that number of workers submitting 0 needs increased with each study as did the number of workers who quit during training (see Table 5).

An additional contributing factor could be the change in incentive structure, from a quota system in study 1 to a piece rate system in studies 2 and 3 (see Table 4). This would be consistent with improved AMT outcomes for quota approaches previously described [64]. There were several rationales for switching structures. One was to smooth the data to create a more uniform distribution because quota structures create bimodal or multimodal distributions. In addition, a payment for every 5 needs seemed to increase confusion and lead to workers entering 5 needs as a single entry. Finally, the need quantities collected in study 1 exceeded expectations, and the benefit of the quota system may not have outweighed the costs given a proficient crowd.

4.6 High Volume of Needs Without Stimulus. Figure 6 reflects that users can readily articulate nearly 50% of the cumulative total need quantity with no official help, independent of evidence that certain types of stimulus can have a significant effect on the count of needs (see Fig. 4). Here, it should be noted that while this figure represents voluntarily selected help specific to the assigned topic area, it is not inclusive of all information that would be useful to workers. Not only did each worker review the instructions and training examples, but a short video summary of instructions was also available, and each worker then saw additional examples during the quiz. Nonetheless, this result shows that these controlled stimuli are beneficial but not required for large quantities of needs.

4.7 User Interface, Quantities, and Rates. The data in Table 6 provide insight into the importance of user interface design and the potential influence on user behavior. In particular, each study recruited approximately the same number of workers. Study 2 and study 3 collected approximately the same number of needs and stories; however, the median duration each worker spent entering needs decreased 45%, from 16 to 11 min. One potential explanation is the effect of interface design. Study 2 was testing a specific treatment effect and did not immediately present workers with a button to end the study while entering needs. This was intentionally withheld until after viewing the assigned help. With this interface, 18 of 347 workers quit without entering any needs.

Study 3 was a modified interface with the rationale that readily available options would be appropriate in a case study application. Here, the button to quit and buttons to access each help were clearly displayed from the beginning of need entry. The number of workers quitting without entering any needs increased to 61 out of 341 (and were paid for completing training). The presence of this greater number of early departures lowered the median duration, but the remaining workers, on average, submitted more needs and the cumulative total was approximately the same. This increase in needs for the nonzero workers could potentially be attributed to the full availability of early and more help.

The increase in maximum duration shown in Table 6 from 116 min to 172 min corresponds to an increase in maximum allowed time for the study (3 hr instead of 2 hr). While a majority of users exit long before this time, a flexible structure allowing engaged users to continue might benefit total counts.

4.8 Limitations and Future Work. A number of limitations to this work should be addressed. Perhaps most important, this data reflect only the total quantity of needs and not the quality or redundancy. Establishing a correlation for finding needs, as has been done for ideation, will require future work, specifically discriminating duplicate and semantically similar needs and then rating unique needs for quality. Duplication from copying might be easily detected, but detecting semantically similar meaning in independently submitted sentences is challenging. This systematic analysis is beyond the scope of the present work. The high quantities and rapidity of needs, readily available crowd participants, and positive effects of stimuli give motivation to pursue this future work. Results of quality ratings would also serve as quantitative outcomes to potentially address barriers facing acceptance of qualitative data described in Sec. 1.6. Research into need quality will also begin to address the composition of crowd expertise discussed in Sec. 4.4. While crowds of nonexperts may provide similarly high need quantities for potentially lower cost, the quality of these needs must be assessed. Future work will also address correlations of quality ratings to level of expertise or experience, for example, if needs submitted by novice users would be rated highest by other novice users.

Second, these results are dependent partly on reputation building and specifics of user interface design. This will have an effect on replicating results within the same crowd, in addition, a strong reputation in one crowd will likely not completely transfer to another when applying this method to long-term application areas such as clinical professionals.

Third, the method relies on what a user says, which may differ from behavior. However, additional validation can come in the form of targeted follow-up observations and a future quality assessment process using a large number of ratings to minimize the influence of individual users.

Finally, the topic areas used in these studies were intentionally very general. The similarity in outcomes for the three general topics is not definitive evidence that the method will be equally effective in a specialized topic, but these studies provide strong rationale and motivation for such future work. Specialized topics will require specialized user groups and may require alternate recruiting strategies, but these groups often already have existing crowd structures potentially available for a large-crowd study. An example in the medical application area would be a national association or annual conference of a medical specialty. The incentive

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

structures devised for this study may not be appropriate for future application areas, but relevant incentives exist for specialized crowds, including peer recognition [65] and formal education credits [66]. A medical case study including assessment of need quality will provide valuable data to further assess these limitations.

5 Conclusions

The combined studies provide strong evidence that users will have the ability to articulate needs directly when the interaction is mediated by sufficient background and instructions, incentives, and stimuli. Furthermore, users can generate nearly half of all needs prior to specific stimuli. The results of these studies indicate that specific stimulus types, such as shared needs and contextual images, can significantly increase the quantity of needs collected, and the incentive structure can be used to influence user behavior. User expertise did not result in a significant difference of needs generated, and although the level of experience was significant, the degree of this change may not warrant exclusive use in practice of high-experience users for need quantity generation. The results indicate that large groups from the general population can effectively generate large quantities of needs, and the results from multiple need topic areas suggest that other topics, including specialized application areas, would be appropriate for future work. The successful, repeated collection of large quantities of needs from large groups warrants further investigation. Additionally, this strongly motivates the development of scalable, automated means of semantically processing and identifying quality needs from large numbers of submitted needs.

Acknowledgment

The Statistical Consulting Service at the University of Minnesota, and in particular Lindsey Dietz, helped with the analysis of this series of experiments. We thank William Durfee, Ph.D. for his valuable input on study design and research goals.

References

[1] Patnaik, D., and Becker, R., 1999, "Needfinding: The Why and How of Uncovering People's Needs," Des. Manage. J., **10**(2), pp. 37–43.
[2] Zenios, S. A., Makower, J., Yock, P. G., Brinton, T. J., Kumar, U. N., Denend, L., and Krummel, T. M., 2010, *Biodesign: The Process of Innovating Medical Technologies*, Cambridge University, New York.
[3] Getzels, J. W., 1979, "Problem Finding: A Theoretical Note," Cognit. Sci., **3**(2), pp. 167–172.
[4] Beckman, S. L., and Barry, M., 2007, "Innovation as a Learning Process: Embedding Design Thinking," Calif. Manage. Rev., **50**(1), pp. 25–56.
[5] Martin, J., Murphy, E., Crowe, J., and Norris, B., 2006, "Capturing User Requirements in Medical Device Development: The Role of Ergonomics," Physiol. Meas., **27**(8), pp. R49–R62.
[6] Cooper, R., and Edgett, S., 2008, "Ideation for Product Innovation: What Are the Best Methods?," PDMA Visions Mag., **1**, pp. 12–17.
[7] Ulrich, K. T., Eppinger, S. D., and Goyal, A., 2011, *Product Design and Development*, Vol. 2, McGraw-Hill, New York.
[8] Kelley, T., and Littman, J., 2001, *The Art of Innovation*, Crown Business, New York.
[9] Gundlin, E., 2000, *The 3M Way to Innovation*, 1st ed., Kodansha Intl., Tokyo.
[10] Alkaya, M., Visser, F. S., and De Lille, C., 2012, "Supporting NPD Teams in Innovation: Structuring User Data on the Foundations of Empathy," Leading Innovation Through Design: Proceedings of the Design Management Institute 2012 International Research Conference, Design Management Institute, pp. 1–8.
[11] Kouprie, M., and Visser, F. S., 2009, "A Framework for Empathy in Design: Stepping Into and Out of the User's Life," J. Eng. Des., **20**(5), pp. 437–448.
[12] Johnson, D. G., Genco, N., Saunders, M. N., Williams, P., Seepersad, C. C., and Hölttä-Otto, K., 2014, "An Experimental Investigation of the Effectiveness of Empathic Experience Design for Innovative Concept Generation," ASME J. Mech. Des., **136**(5), p. 051009.
[13] Elliott, R., and Jankel-Elliott, N., 2003, "Using Ethnography in Strategic Consumer Research," Qual. Mark. Res., **6**(4), pp. 215–223.
[14] Savenye, W., and Robinson, R., 1996, "Qualitative Research Issues and Methods: An Introduction for Educational Technologists," *Handbook of Research for Educational Communications and Technology*, Macmillan Library Reference, New York, pp. 1171–1195.
[15] Weiss, R., 1995, *Learning From Strangers: The Art and Method of Qualitative Interview Studies*, 1st ed., Free Press, New York.
[16] Rubin, H. J., and Rubin, I. S., 2011, *Qualitative Interviewing: The Art of Hearing Data*, SAGE Publications, Thousand Oaks.
[17] Griffin, A., and Hauser, J. R., 1993, "The Voice of the Customer," Mark. Sci., **12**(1), pp. 1–27.
[18] Bayus, B. L., 2008, "Understanding Customer Needs," *The Handbook of Technology and Innovation Management*, S. Shane, ed., Wiley, West Sussex, UK, pp. 115–141.
[19] Goldenberg, J., Lehmann, D. R., and Mazursky, D., 2001, "The Idea Itself and the Circumstances of Its Emergence as Predictors of New Product Success," Manage. Sci., **47**(1), pp. 69–84.
[20] Balachandra, R., and Friar, J. H., 1997, "Factors for Success in R&D Projects and New Product Innovation: A Contextual Framework," IEEE Trans. Eng. Manage., **44**(3), pp. 276–287.
[21] Holbrook, M. B., and Hirschman, E. C., 1982, "The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun," J. Consum. Res., **9**(2), pp. 132–140.
[22] Babin, B. J., Darden, W. R., and Griffin, M., 1994, "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value," J. Consum. Res., **20**(4), pp. 644–656.
[23] Desmet, P. M., and Hekkert, P., 2007, "Framework of Product Experience," Int. J. Des., **1**(1), pp. 57–66.
[24] Kudrowitz, B. M., and Wallace, D., 2013, "Assessing the Quality of Ideas From Prolific, Early-Stage Product Ideation," J. Eng. Des., **24**(2), pp. 120–139.
[25] Dean, D. L., Hender, J. M., Rodgers, T. L., and Santanen, E. L., 2006, "Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation," J. Assoc. Inf. Syst., **7**(10), pp. 646–698.
[26] Osborn, A. F., 1953, *Applied Imagination*, Scribner's, New York.
[27] Nunamaker, J. F., Jr., Briggs, R. O., Mittleman, D. D., Vogel, D. R., and Balthazard, P. A., 1996, "Lessons From a Dozen Years of Group Support Systems Research: A Discussion of Lab and Field Findings," J. Manage. Inf. Syst., **13**(3), pp. 163–207.
[28] Martins, L. L., Gilson, L. L., and Maynard, M. T., 2004, "Virtual Teams: What do We Know and Where Do We Go From Here?," J. Manage., **30**(6), pp. 805–835.
[29] VanGundy, A. B., 1984, "Brain Writing for New Product Ideas: An Alternative to Brainstorming," J. Consum. Mark., **1**(2), pp. 67–74.
[30] Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., and Nunamaker, J. F., 1992, "Electronic Brainstorming and Group Size," Acad. Manage. J., **35**(2), pp. 350–369.
[31] Pinsonneault, A., Barki, H., Gallupe, R. B., and Hoppen, N., 1999, "Electronic Brainstorming: The Illusion of Productivity," Inf. Syst. Res., **10**(2), pp. 110–133.
[32] DeRosa, D. M., Smith, C. L., and Hantula, D. A., 2007, "The Medium Matters: Mining the Long-Promised Merit of Group Interaction in Creative Idea Generation Tasks in a Meta-Analysis of the Electronic Group Brainstorming Literature," Comput. Hum. Behav., **23**(3), pp. 1549–1581.
[33] Paulus, P. B., Kohn, N. W., and Arditti, L. E., 2011, "Effects of Quantity and Quality Instructions on Brainstorming," J. Creat. Behav., **45**(1), pp. 38–46.
[34] Diehl, M., and Stroebe, W., 1987, "Productivity Loss in Brainstorming Groups: Toward the Solution of a Riddle," J. Pers. Soc. Psychol., **53**(3), pp. 497–509.
[35] Dugosh, K. L., and Paulus, P. B., 2005, "Cognitive and Social Comparison Processes in Brainstorming," J. Exp. Soc. Psychol., **41**(3), pp. 313–320.
[36] Morgan, J., and Wang, R., 2010, "Tournaments for Ideas," Calif. Manage. Rev., **52**(2), pp. 77–97.
[37] Jansson, D. G., and Smith, S. M., 1991, "Design Fixation," Des. Stud., **12**(1), pp. 3–11.
[38] Linsey, J., Tseng, I., Fu, K., Cagan, J., Wood, K., and Schunn, C., 2010, "A Study of Design Fixation, Its Mitigation and Perception in Engineering Design Faculty," ASME J. Mech. Des., **132**(4), p. 041003.
[39] Viswanathan, V. K., and Linsey, J. S., 2013, "Design Fixation and Its Mitigation: A Study on the Role of Expertise," ASME J. Mech. Des., **135**(5), p. 051008.
[40] IDEO, 2003, Method Cards: 51 Ways to Inspire Design, Palo Alto, CA.
[41] Lin, J., and Seepersad, C. C., 2007, "Empathic Lead Users: The Effects of Extraordinary User Experiences on Customer Needs Analysis and Product Redesign," ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. 289–296.
[42] Chesbrough, H. W., 2006, "The Era of Open Innovation," Managing Innovation Change, **127**(3), pp. 34–41.
[43] Dow, S., Gerber, E., and Wong, A., 2013, "A Pilot Study of Using Crowds in the Classroom," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'13, ACM, pp. 227–236.
[44] Stone, T., and Choi, S.-K., 2013, "Extracting Consumer Preference From User-Generated Content Sources Using Classification," ASME Paper No. DETC2013-13228.
[45] Faste, H., 2011, "Opening 'Open' Innovation," Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces, DPPI'11, ACM, pp. 54:1–54:8.
[46] Nijstad, B. A., Stroebe, W., and Lodewijkx, H. F., 2002, "Cognitive Stimulation and Interference in Groups: Exposure Effects in an Idea Generation Task," J. Exp. Soc. Psychol., **38**(6), pp. 535–544.
[47] Dugosh, K. L., Paulus, P. B., Roland, E. J., and Yang, H.-C., 2000, "Cognitive Stimulation in Brainstorming," J. Pers. Soc. Psychol., **79**(5), pp. 722–735.
[48] Kohn, N. W., Paulus, P. B., and Choi, Y., 2011, "Building on the Ideas of Others: An Examination of the Idea Combination Process," J. Exp. Soc. Psychol., **47**(3), pp. 554–561.

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use

[49] Dennis, A. R., Minas, R. K., and Bhagwatwar, A. P., 2013, "Sparking Creativity: Improving Electronic Brainstorming With Individual Cognitive Priming," J. Manage. Inf. Syst., **29**(4), pp. 195–216.
[50] Rosenthal, S., and Capper, M., 2006, "Ethnographies in the Front End: Designing for Enhanced Customer Experiences," J. Prod. Innovation Manage., **23**(3), pp. 215–237.
[51] Toh, C. A., and Miller, S. R., 2014, "The Impact of Example Modality and Physical Interactions on Design Creativity," ASME J. Mech. Des., **136**(9), p. 091004.
[52] Martin, J., and Barnett, J., 2012, "Integrating the Results of User Research Into Medical Device Development: Insights From a Case Study," BMC Med. Inf. Decis. Making, **12**(74), pp. 1–10.
[53] Kramer, D. B., Xu, S., and Kesselheim, A. S., 2012, "How Does Medical Device Regulation Perform in the United States and the European Union? A Systematic Review," PLoS Med., **9**(7), p. 1001276.
[54] Money, A., Barnett, J., Kuljis, J., Craven, M., Martin, J., and Young, T., 2011, "The Role of the User Within the Medical Device Design and Development Process: Medical Device Manufacturers' Perspectives," BMC Med. Inf. Decis. Making, **11**(15), pp. 1–12.
[55] Shah, A., and Alshawi, S., 2010, "The Role of User Requirements Research in Medical Device Development," European and Mediterranean Conference on Information Systems, Abu Dhabi, UAE, EMCIS2010, pp. 1–25.
[56] Poulton, B., 1999, "User Involvement in Identifying Health Needs and Shaping and Evaluating Services: Is it Being Realised?," J. Adv. Nurs., **30**(6), pp. 1289–1296.
[57] Ram, M., Campling, N., Grocott, P., and Weir, H., 2008, "A Methodology for a Structured Survey of the Healthcare Literature Related to Medical Device Users," Evaluation, **14**(1), pp. 49–73.
[58] Vredenburg, K., Mao, J., Smith, P., and Carey, T., 2002, "A Survey of User-Centered Design Practice," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves, ACM, pp. 471–478.
[59] Aquino Shluzas, L., Steinert, M., and Leifer, L. J., 2011, "Designing to Maximize Value for Multiple Stakeholders: A Challenge to Med-Tech Innovation," 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 10: Design Methods and Tools pt. 2, Lyngby/Copenhagen, Aug. 15–19, pp. 159–166.
[60] Shah, S., and Robinson, I., 2007, "Benefits of and Barriers to Involving Users in Medical Device Technology Development and Evaluation," Int. J. Technol. Assess. Health Care, **23**(1), pp. 131–137.
[61] Kauppinen, M., Kujala, S., Aaltio, T., and Lehtola, L., 2002, "Introducing Requirements Engineering: How to Make a Cultural Change Happen in Practice," IEEE Joint International Conference on Requirements Engineering, pp. 43–51.
[62] Paolacci, G., Chandler, J., and Ipeirotis, P., 2010, "Running Experiments on Amazon Mechanical Turk," Judgment Decis. Making, **5**(5), pp. 411–419.
[63] Bretz, F., Hothorn, T., and Westfall, P., 2010, *Multiple Comparisons Using R*, CRC Press, Boca Raton, FL.
[64] Mason, W., and Watts, D. J., 2010, "Financial Incentives and the Performance of Crowds," ACM SigKDD Explor. Newslett., **11**(2), pp. 100–108.
[65] Nov, O., Anderson, D., and Arazy, O., 2010, "Volunteer Computing: A Model of the Factors Determining Contribution to Community-Based Scientific Research," Proceedings of the 19th International Conference on World Wide Web, ACM, pp. 741–750.
[66] Curran, V. R., and Fleet, L., 2005, "A Review of Evaluation Outcomes of Web-Based Continuing Medical Education," Med. Educ., **39**(6), pp. 561–567.

Downloaded From: https://mechanicaldesign.asmedigitalcollection.asme.org/ on 11/05/2015 Terms of Use: http://www.asme.org/about-asme/terms-of-use