The Effect of Peripheral Micro-tasks on Crowd Ideation

Victor Girotto[1] **Erin Walker**[1]
[1] School of Computing, Informatics,
and Decision Systems Engineering
Arizona State University
Tempe, AZ, USA
{victor.girotto, erin.a.walker}@asu.edu

Winslow Burleson[2]
[2] Rory Meyers College of Nursing
New York University
New York, NY, USA
wb50@nyu.edu

ABSTRACT

Research has explored different ways of improving crowd ideation, such as presenting examples or employing facilitators. While such support is usually generated through peripheral tasks delegated to crowd workers who are not part of the ideation, it is possible that the ideators themselves could benefit from the extra thought involved in doing them. Therefore, we iterate over an ideation system in which ideators can perform one of three peripheral tasks (rating originality and usefulness, similarity, or idea combination) on demand. In controlled experiments with workers on Mechanical Turk, we compare the effects of these secondary tasks to simple idea exposure or no support at all, examining usage of the inspirations, fluency, breadth, and depth of ideas generated. We find tasks to be as good or better than exposure, although this depends on the period of ideation and the fluency level. We also discuss implications of inspiration size, homogeneity, and frequency.

Author Keywords

Crowdsourcing; ideation; creativity; microtasks.

ACM Classification Keywords

H.5.3. Group and Organization Interfaces: Computer-supported cooperative work.

INTRODUCTION

With the advent of crowdsourcing, people can now collectively accomplish a wide range of tasks that could not otherwise be done by a single human or computer. One approach to crowdsourcing that has stood out is the use of micro-task markets such as Amazon's Mechanical Turk (MTurk) [14]. In this approach, many workers perform small tasks that together approximate the quality of experts. Using micro-task markets, researchers have been able to achieve good results on a wide variety of tasks [2,5].

In our work, we leverage a similar micro-tasks paradigm to achieve creative solutions in response to complex problems. Creativity thrives on diversity and exploration. It is about creating something that is both novel, breaking away from common knowledge or practices, but at the same time being appropriate or useful [11]. From designing T-shirts (www.threadless.com) to solving tough technical challenges (www.innocentive.com), there are many examples of the crowd performing tasks that rely on their creativity.

Why explore the creativity of the crowds? The first reason is that a great number of people will generate a great number of ideas. Furthermore, the heterogeneity of the crowd can increase the potential of ideas being sparked that otherwise wouldn't [8]. However, there are also issues that need to be carefully considered in a system that tries to tap into the crowd's creativity. Issues such as cognitive interference or social loafing can increase together with the number of ideators [8]. Therefore, crowd ideation needs to be carefully designed in order to improve, not hinder the creative output.

A popular method used for generating ideas is typically brainstorming, which seeks to increase the number of ideas generated by encouraging intensive exploration of ideas while restricting criticism [25]. In the crowd context, just like in smaller groups, people have tried to enhance idea generation during brainstorming sessions in different ways, many times employing other individuals or workers, outside of ideation, to do tasks whose output will benefit ideators. We call these tasks *peripheral tasks*. The result of their work is then presented in some way to crowd ideators. For example: Yu et al. [31] had workers generate problem schemas, and subsequently used them to enhance ideation performance of other workers in a subsequent study. However, the extra cognitive effort that is required to perform these tasks could potentially benefit ideators as much as just using their results.

This paper, therefore, examines the effect that performing peripheral tasks has on ideation. More specifically, it embeds three types of peripheral tasks—rating, similarity, and combination—into an online brainstorming session. We explore the following questions:

1. How does performing peripheral micro-tasks affect ideation performance?
2. Do different types of peripheral tasks affect ideation differently? If so, how?

CHI 2017, May 06-11, 2017, Denver, CO, USA
© 2017 ACM. ISBN 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025464

Exploration of these questions could allow ideation systems to move from passive to active forms of inspiration and support, resulting in more data collection during an ideation session, aiding in convergent tasks such as idea selection. A similar approach has been explored by Siangliulue and colleagues through the IdeaHound system [28]. IdeaHound allows users to physically cluster semantically related ideas together in a virtual workspace. This organization enables the system to infer a semantic model of the ideas. Our approach differs in that it makes this data collection explicit rather than implicitly building it in the UI interactions of the system. In other words, rather than inferring semantic relatedness by examining how ideators cluster ideas together, we explicitly ask them to judge the similarity of two ideas. Our focus, however, is on how doing these tasks affects ideation performance, rather than examining their result.

In the remainder of this paper, we review literature related to creativity and crowdsourcing. We then describe a system built to allow ideators to perform small tasks during ideation, and describe metrics for its evaluation, including a tree-based representation of individual ideators' performance. We then describe four experiments, evaluating how their combined results answer the questions above. Generally, we find that tasks are just as useful as simple idea exposure, with rating and combination tasks even outperforming it in certain situations. We also explore how inspiration size, frequency, and homogeneity affects ideation.

RELATED WORK

Creativity: Convergent and Divergent Processes

Creativity can be defined, at the most basic level, as the production of something original and appropriate [11]. While there are many different theories of the underlying nature of creative processes and products [17], our specific interest is in the dichotomy between divergent and convergent processes. Divergent processes are those that generate a wide variety of ideas, thus increasing the solution space [6,17]. On the other hand, convergent processes are those that involve the selection of a particular number of the best ideas, seeking to reduce ambiguity and the size of the possible solution space [6,17]. Both processes are necessary for creativity: generating variability (divergence) without effectively exploring and evaluating your ideas (convergence) can lead to lost opportunities or disastrous changes [6].

Brainstorming

Much of the effort in research and practice in improving creativity has focused on supporting divergent processes through brainstorming. Brainstorming was popularized by Osborn in the 1950's, and consists of a few simple rules, such as holding back on criticism and building on the ideas of others [25]. There are in fact benefits to this approach, as attending to the ideas of others can be inspiring [19,23]. But it is vulnerable to factors such as evaluation apprehension, free riding, or perhaps more influentially, production blocking—that is, not being able to share or generate new ideas while someone is sharing theirs [9].

By moving from interactive co-located groups to electronic communication media, production blocking and other issues can be lessened [7]. For example, individuals in an ideation group that uses an instant messaging channel for collaboration don't have to wait their turn to speak, and can choose to attend to others' ideas as they desire. Furthermore, as communication technologies advance and the world grows increasingly connected, ever increasing group sizes become more feasible and the possibility of synergy between the participants' ideas can also increase [7,8]. In other words, by being exposed to more ideas, it becomes more likely that users will see concepts that may spark new ideas.

One common feature of group brainstorming sessions is to hear ideas developed by others and build on them. Exposure to other ideas may have different effects depending on the type of exposure, with the possibility of either stimulating or hindering creativity. On the one hand, exposure to a diverse set of ideas can increase breadth of ideation, while exposure to a homogenous set of ideas can increase the depth of exploration within each semantic category [24]. On the other hand, exposure to ideas might lead to conformity and fixation effects, where ideators take the main concepts of the ideas they were exposed to and use them in their solutions [13,19,30]. This cognitive interference may affect the exploration of the solution space, causing answers to be more like each other. There is evidence that the nature of the external influence as well as *how* you attend to it defines its effect on you. Paying too much attention to the superficial details of the example ideas may lead to fixation [13,19], while a higher-level view of ideas, possibly in the form of analogies or schemas, can improve idea generation [31,32].

Perhaps the clearest way to understand the effect of external stimuli comes from the Search for Ideas in Associate Memory (SIAM) model [23,24]. This model assumes the existence of two kinds of memory: a low-capacity short-term memory, i.e. working memory (WM) and a high-capacity long-term memory (LTM). It proposes that idea generation involves these two memories in two stages: first, an ideator retrieves a concept along with its features from LTM into the WM (e.g. the concept *hotel* has the feature "has rooms" [23]), and then generates ideas based it. When the ideator continuously fails at generating new ideas based on the current concept, he or she may activate another concept from the search queue and try again. This queue is comprised of items such as the problem definition or previously generated ideas. When an idea is shown to an ideator, it can be added to the search queue if it is sufficiently attended to [24]. Therefore, it can be said that an ideator had a great *breadth* of ideation if he or she explored many concepts and great *depth* if he or she developed many ideas within a concept.

Peripheral Ideation Tasks

Much of the support for crowd ideation sessions comes from peripheral tasks done by other MTurk crowd workers. These tasks can be classified along four main categories: rating, combination, inspiration design, and problem abstraction. In crowd ideation contexts, workers outside of the ideation are typically tasked with performing them. However, we argue that these tasks have promise for improving the quality of brainstorming when executed by the ideators themselves. We now review the rating and combination tasks, which are explored in depth in this paper. They were chosen due to their simplicity and for their usage of external ideas, allowing for a cleaner comparison with simple idea exposure. We leave the remaining two categories for future work.

Rating

Rating an idea or artifact on one or more dimensions is one of the most common forms of peripheral tasks. We are concerned with two types of rating: the nature of the idea (originality and usefulness), or its similarity to other ideas. Rating can support ideation by, for example, presenting the most creative or diverse sets of ideas as inspiration to other ideators [24]. Comparisons are either relative—rate ideas in terms of each other—or absolute—rate ideas individually on a given scale. For example, Siangliulue et al. [27] looked into rating ideas based on their similarity in two different ways: for the first, they presented workers with three ideas, asking them to choose which of two ideas was more similar to another. They also tasked workers with rating 30 pairs of ideas on their similarity, using a 7-point scale. Other ways ratings tasks have been performed are to ask workers to rate ideas in terms of novelty and quality (or similar dimension), both on a 7-point scale [4,28,32].

What could be the effect of performing rating tasks on an ideator's performance? By rating others' ideas, ideators would be exposed to a small number of raw ideas which, as we have already reviewed, may produce either stimulation or fixation effects. These tasks would require users to think critically about the examples in order to rate them in terms of their similarity, usefulness, or originality. Attending to the stimuli provided by external ideas is a requirement for the idea to influence ideation [10]. In fact, the SIAM model proposes that an idea will only be added to the search cue if sufficiently attended to [23]. Therefore, having users actively engage in analyzing other ideas may promote a larger effect—either of stimulation or interference—than if just passively being exposed to them. But this may depend on the nature of the idea and how it is displayed [24,29], as rating others' ideas may positively affect the creativity, diversity, depth, and quantity of ideas generated.

A negative effect may come from engaging in criticism or judgment during brainstorming, which usually discourages criticism as it may lead to evaluation apprehension. However, there is evidence that this effect does not account for most productivity loss in brainstorming [9]. In fact, criticism has even been found to improve performance [21]. Furthermore, while rating an idea's originality and usefulness may produce such effect, other rating tasks such as rating the similarity may not yield such effects due to their non-judgmental nature.

Combination

Combination tasks revolve around combining characteristics of two ideas, with the goal of generating a third idea. Combination engages divergent processes that can produce better, more diverse ideas [16]. In a study by Yu & Nickerson [33], for example, workers were presented with two different ideas, generated by other workers, for the design of a chair. Their task was to design a new chair that combined aspects of the two other ideas. The resulting design was then passed to a different set of workers for further processing. Combinations could also employ higher representations of ideas—schemas—as explored by Yu, Kittur, & Kraut, who used them in both exploration and generation of ideas [31,32]. Schemas can facilitate analogical reasoning when combining ideas, focusing ideators on the core principles of an idea rather than on its surface features.

While, to our knowledge, there are no other examples of combination tasks in crowdsourced ideation within a microtask platform such as MTurk, the microtasks literature provides other instances of similar tasks, which demonstrate the feasibility of assigning workers to combine two or more different objects. On a general level, the CrowdForge framework [15] defines three types of tasks: partition, in which tasks are broken up in smaller pieces; map, in which a task is processed by one or more workers; and reduce, in which the result of multiple worker's tasks are merged into a single output. One of their examples demonstrates how workers could write an article by breaking it down into an outline, assigning workers to write down facts for each topic in an outline, and having workers merge those facts into paragraphs. In the case of ideation tasks, the merge step could be defined as combining two ideas. More specifically, however, combination of results from different workers is a common feature in MTurk workflows, generally for quality assurance purposes (e.g. [2,5]).

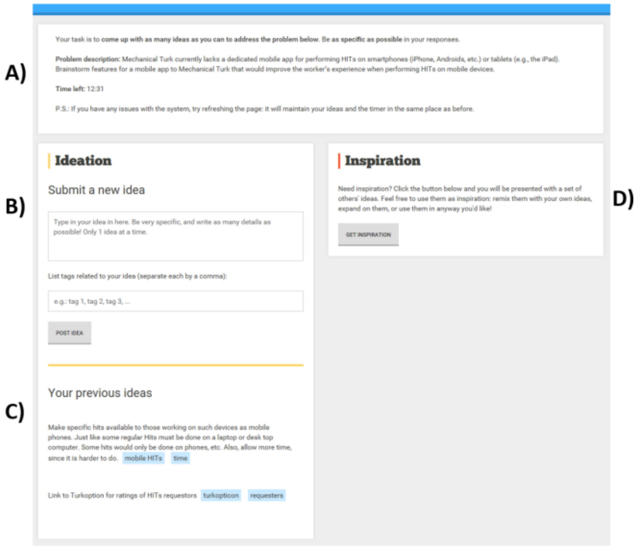
Thus, our hypothesized effect of combination tasks on ideation relates to that of rating: users are exposed to new ideas and have to attend to them to perform the combination. However, unlike the rating tasks, this is not an entirely convergent process, but also includes a divergent step of generating a new idea [16]. This may counterbalance any fixation that can happen while attending to the ideas that are to be combined. On the other hand, the depth of the produced ideas might decrease due to this increased divergence.

SYSTEM AND WORKFLOW

We developed an online ideation system that enables the creation of timed asynchronous ideation sessions, as well as a mechanism for seeing other people's ideas upon request via an inspiration button, thus allowing ideators to *pull* inspirations whenever they choose to do so. This is in line with previous approaches (e.g. [4,29]). An alternative to this approach would be to *push* inspirations at regular time intervals. This would ensure that every ideator was exposed to the same number of inspirations, allowing a clearer comparison of the effect of the different types of tasks.

However, one of our goals was to see if embedding tasks into inspirations would detract from users' interest in using the inspiration mechanism or decrease performance. A push approach would hinder us from exploring this. Furthermore, the SIAM model predicts that pushing inspirations could negatively affect performance [22], since it could interrupt a users' train of thought. In fact, Siangliulue et al. found issues with fluency using a push approach [29]. Therefore, we allowed users to request inspirations on demand.

The system is comprised of four main parts (Figure 1). Although the figure depicts the system in its final iteration, its overall structure as described in this section was maintained throughout the sessions, with a few incremental differences that will be pointed out in each experiment's section. At the top (A), the system displays instructions, the problem definition, and a timer. On the left is the ideation panel. It consists of a form for entering an idea along with a list of user defined tags associated with it (B), and a list of the user's previously submitted ideas and tags (C). On the right side is the inspiration panel (D). When the button is clicked, the user is presented with a set of ideas and depending on their condition, a task associated with them. This mechanism draws randomly from a pool of ideas generated in previous experiments.



Your task is to **come up with as many ideas as you can to address the problem below**. Be **as specific as possible** in your responses.

Problem description: Mechanical Turk currently lacks a dedicated mobile app for performing HITs on smartphones (iPhone, Androids, etc.) or tablets (e.g., the iPad). Brainstorm features for a mobile app to Mechanical Turk that would improve the worker's experience when performing HITs on mobile devices.

Time left: 12:31

P.S.: If you have any issues with the system, try refreshing the page: it will maintain your ideas and the timer in the same place as before.

Submit a new idea

Type in your idea in here. Be very specific, and write as many details as possible! Only 1 idea at a time.

List tags related to your idea (separate each by a comma):

e.g.: tag 1, tag 2, tag 3, ...

Your previous ideas

Make specific hits available to those working on such devices as mobile phones. Just like some regular Hits must be done on a laptop or desk top computer. Some hits would only be done on phones, etc. Also, allow more time, since it is harder to do. mobile HITs time

Link to Turkoption for ratings of HITs requestors turkopticon requesters

Need inspiration? Click the button below and you will be presented with a set of others' ideas. Feel free to use them as inspiration: remix them with your own ideas, expand on them, or use them in anyway you'd like!

Figure 1. Screenshot of the ideation system as used in the final experiment, comprised of the following parts: A) Problem description and timer; B) idea submission input; C) list of the users' submitted ideas; D) inspiration panel.

When users access the system, they first see a page stating how much time the session lasts, and asking them to move forward only if absolutely sure that they can commit their full attention for the specified amount of time. Following that, users would see another page describing the system, including the inspiration mechanism (if any), and how to use it. Upon finishing the instructions, users begin the ideation session. After the timer is done, the system presents users with a thank you message, a user ID (used for payment), and a link to a short post-session survey.

For every study in this paper, the problem that ideators were tasked to ideate on was: "*Mechanical Turk currently lacks a dedicated mobile app for performing HITs on smartphones (iPhone, Androids, etc.) or tablets (e.g., the iPad). Brainstorm N features for a mobile app to Mechanical Turk that would improve the worker's experience when performing HITs on mobile devices. Be as specific as possible in your responses.*" This task, suggested by Krynicki [18], was chosen because it has been successfully used in previous studies [3,18] and MTurk users have knowledge about the issue and may be motivated to contribute to it, as it could increase their opportunities for engaging with HITs and improving their income. Both motivation and knowledge are key to creativity [1].

METRICS

For each study, we report the following metrics:

- *Fluency*: number of ideas generated by the user.
- *Number of inspirations*: number of times the user clicked the inspiration button.
- *Inspiration influence*: a user's average similarity between an idea and the most similar of its preceding inspirations.

More central to our interests, however, are metrics of *breadth* and *depth*, which we extracted from an *ideation tree*, described below. Tree representations have been previously suggested to measure or visualize ideation outcome [12,20], and the semantics of the different branches of a tree can reflect the usual discrete categorization of ideas traditionally used in creativity research [26], while their depth can represent the notion of ideation within one category [23]. This tree is built from a chronological list of user actions—they either add a new idea or request an inspiration. In the tree, similarity between ideas is measured using Latent Semantic Analysis (LSA) [3]. For this paper, our LSA corpus was built on 5640 ideas generated to solve the same problem that we explore in this paper. This corpus comes both from our own studies (2115 ideas) and the corpus shared by the authors of [4] (3525 ideas). Figure 2 shows the tree and idea pool in five different points in time during a user's ideation:

1. We add the first user idea as the child of a dummy node;
2. For the second user generated idea, we compare it either with every node that is already in the tree, or with every inspiration previously seen. If the LSA similarity to any of those is greater than a given threshold (we used 0.5), we add it as a child of the most similar node. In this point in time, idea 2 was most similar to idea 1, and is added as its child;
3. At the third point in time, the user has generated a third idea. Again, we compare it to every node already in the tree. In this case, none of the similarities exceeded the threshold, so this idea is added as a new child of the root node, representing an estimated new category of ideation.

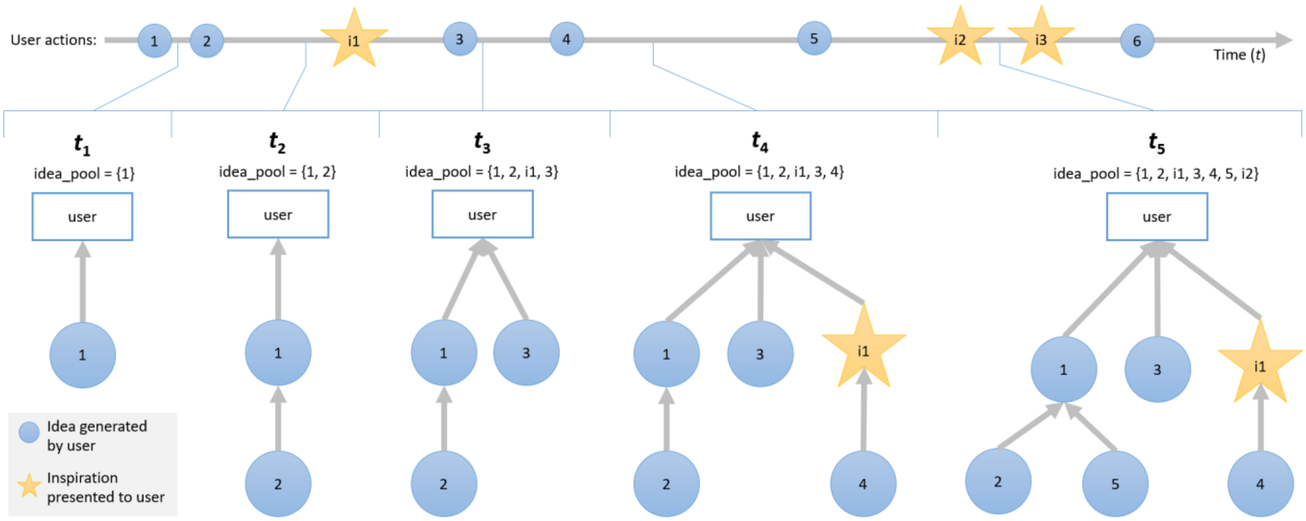


Figure 2. Snapshots of the ideation tree and idea pool in five different points of time in a given user's ideation.

Also note that between t_2 and t_3 the ideator has requested an inspiration, which is added to the idea pool but not to the tree;

4. The user's 4th idea is compared to every node in the idea pool. In this case, a previously seen inspiration is the most similar. Therefore, we add it as a child of the inspiration node, which we then add as a new child of the root node;
5. Finally, another idea has been generated by the user. It is compared to previous ideas and inspirations, and is added as a child to the most similar node, which is the first one.

From this tree, we extract the two metrics:

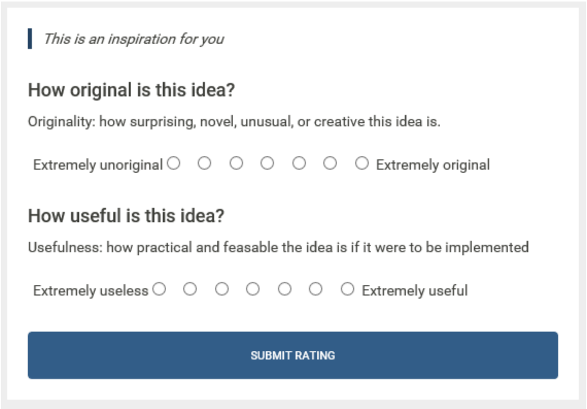
- *Breadth*: the number of children in the root node. These were the ideas that, at the time they were added, were not similar enough to be considered a continuation of another idea, therefore creating a new branch of ideas. For example, in Figure 2 at t_5, the breadth would be 3.
- *Depth*: the number of nodes in the branch with the most number of nodes. For example, in Figure 2 at t_5, the depth would be 3.

As a check on this measure, we additionally calculated the metrics described by Chan et al. [4], also built using Latent Semantic Analysis (LSA). In [4], breadth was the mean distance between each pair of ideas generated by a user. Depth was the maximum similarity between the ideas generated by a user. Our metrics are significantly correlated with these, at $r = 0.650$, $p < 0.001$ for breadth, and $r = 0.564$, $p = 0.001$ for depth. And while there may be concerns relative to the metric's sensitivity to the threshold value, we have found that threshold changes in either direction do not result in drastic changes in the results. For example, changing it to 0.7 resulted in a mean difference of 2.3 (SD=2.2) in breadth. Therefore, we believe that this metric is both valid and capable of more accurately representing the notions of breadth and depth.

EXPERIMENT 1: RATING TASK

For the first experiment, we chose to start our exploration with rating tasks. Due to their simplicity, they can easily and quickly be done by any worker, and they are very effective for supporting convergence processes. We used an earlier version of the system than the one described in the system section, which differed as follows: the input box for the idea was on the top panel, along with the problem description; users did not have to input tags for their ideas; Lastly, the inspiration box did not have any instructions regarding how inspirations could be used. This study had three conditions:

1. **Control**: This condition is equivalent to a nominal group in typical brainstorming settings. There is no inspiration panel, and thus no external stimulus. Users type their ideas, and can see the list of their own ideas.
2. **Exposure**: In this condition, the inspiration panel is visible. When the inspiration button is clicked, it displays one idea from the pool of past ideas, without any task associated with it. The idea disappears when the user clicks the "done" button.
3. **Rating**: This condition is similar to the exposure condition. However, when the inspiration button is clicked, in addition to the idea, users also received a task prompting them to rate the inspiration idea in 2 dimensions: originality and practicality (Figure 3). After submitting the rating, the idea disappears.



This is an inspiration for you

How original is this idea?

Originality: how surprising, novel, unusual, or creative this idea is.

Extremely unoriginal ○ ○ ○ ○ ○ ○ ○ Extremely original

How useful is this idea?

Usefulness: how practical and feasable the idea is if it were to be implemented

Extremely useless ○ ○ ○ ○ ○ ○ ○ Extremely useful

Figure 3. Rating task interface.

We published a MTurk HIT that directed workers to our system. 60 workers participated in this study (at least 1000 completed HITs, approval > 95%, US only), but 1 (exposure condition) was excluded from the analysis due to an abnormal number of inspirations requested (142). In total, 559 ideas were generated. Each worker ideated for 18 minutes, filled out a small survey at the end of the session, and was compensated $2. Workers also received an ideation qualification on MTurk (awarded after every experiment). Subsequent experiments reported here required workers to not have this qualification, thus ensuring participants were unique for each session.

Results

Condition	Workers	Ideas / Worker	Insp. / Worker
Baseline	19	10.37 (4.16)	-
Exposure	19	9.53 (4.62)	12.16 (10.64)
Rating	21	8.62 (5.02)	6.67 (6.80)

Table 1. Fluency and inspiration metrics for experiment.

Tables 1 and 2 summarize the metrics across the different conditions. A one-way ANOVA test shows no significant difference in fluency, F(2,56) = 0.713, p = 0.495. There was a marginally significant difference in number of inspirations requested across the two conditions, F(1,38) = 3.855, p = 0.057. We do, however, find a difference in inspiration influence between the exposure and rating conditions, F(1,38) = 9.855, p = 0.003.

Condition	Breadth	Depth	Influence
Baseline	5.37 (2.54)	4.74 (3.69)	-
Exposure	6.40 (3.20)	3.25 (1.74)	0.23 (0.90)
Rating	5.10 (2.70)	3.29 (1.52)	0.12 (0.12)

Table 2. Breadth, depth, and influence for experiment 1.

We calculated a Mixed Generalized Linear Model (GLM) with breadth as outcome variable, condition as factor, and the fluency as covariate. We included the interaction between condition and fluency in the model. We found a marginally significant interaction effect between condition and fluency, F(2, 56) = 3.078, p = 0.054, and no main effect of condition on breadth, F(2,56) = 1.374, p = 0.262.

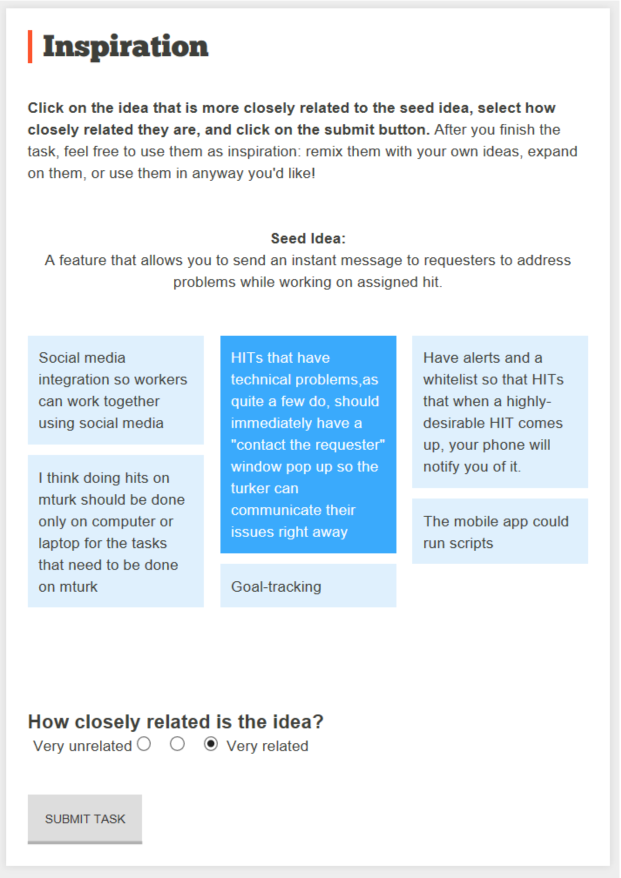
As the depth of user ideas followed a negative binomial distribution rather than a normal distribution, we conducted a negative binomial regression with depth as outcome, condition as factor, and fluency as covariate. The interaction between fluency and condition was included in the model. We found no significant interaction effect or main effect of condition on depth, Wald Chi-Square = 4.099, p = 0.129.

Discussion for Study 1

While we see only a marginal effect of the interaction between condition and fluency on breadth, we find no clear advantage in any condition. The fact that breadth seemed to be more affected than depth may spring from inspirations being randomly drawn from the pool of ideas, which will likely create a heterogeneous set of examples. Past work has shown that a heterogeneous set of examples will improve diversity of ideas [24,27]. Having no clear advantage could mean a problem either in the intervention (e.g. it is too simple) or in how users performed it (e.g. they did not attend to it). There may also have been confusion on how users should use the ideas in the rating task. For example, a user declared feeling that the inspiration they got would invalidate using that idea: "*I think it hindered me more than it helped because it just provided an example that I then couldn't use*". Perhaps guidelines might be effective in helping users better use the inspirations.

EXPERIMENTS 2A & 2B: SIMILARITY CHOICE TASK

In experiment 1, we found no clear advantages over the baseline, even though there was a larger influence in the exposure condition. Given these results, we decided to change the task to similarity comparison, the number of ideas displayed, and to add a clarification on how they could use the inspirations (see the text at the top of Figure 4).



Inspiration

Click on the idea that is more closely related to the seed idea, select how closely related they are, and click on the submit button. After you finish the task, feel free to use them as inspiration: remix them with your own ideas, expand on them, or use them in anyway you'd like!

Seed Idea:
A feature that allows you to send an instant message to requesters to address problems while working on assigned hit.

Social media integration so workers can work together using social media

I think doing hits on mturk should be done only on computer or laptop for the tasks that need to be done on mturk

HITs that have technical problems,as quite a few do, should immediately have a "contact the requester" window pop up so the turker can communicate their issues right away

Goal-tracking

Have alerts and a whitelist so that HITs that when a highly-desirable HIT comes up, your phone will notify you of it.

The mobile app could run scripts

How closely related is the idea?

Very unrelated ○ ○ ● Very related

Figure 4. The task panel for condition 3. Users were shown a seed idea along with 6 other ideas, and were asked to click on the most similar one (in this case, the user clicked on the dark blue idea), as well as rating their degree of relationship.

Experiment 2A

In this experiment, the task condition presents the user with one seed idea along with 6 other ideas, asking him or her to choose the most similar to the seed (Figure 4). The number of ideas was chosen to maximize the possibility of similar ideas being shown, as well as to explore the result of a more dramatic increase in the number of ideas shown per inspiration. We expected this to yield a stronger influence on ideators' breadth, as they would be exposed to more ideas. We also hypothesized that similarity comparisons would force to user to think more abstractly about the ideas in order to find common features between them, thus reducing fixation and possibly improving breadth.

This second experiment followed the same method as the first, with the two key differences above. 60 workers participated in this study (at least 1000 completed HITs, approval > 95%, US only). In total, 492 ideas were generated. Each worker ideated for 18 minutes, filled out a small survey at the end of the session, and was paid $2.

Results

Condition	Workers	Ideas / Worker	Insp. / Worker
Baseline	20	7.45 (5.51)	-
Exposure	22	8.50 (4.34)	2.00 (2.19)
Similarity	18	8.61 (2.87)	2.28 (1.77)

Table 3. Fluency and inspiration metrics for experiment 2A.

Tables 3 and 4 summarize the metrics for this experiment. A one-way ANOVA test shows no difference in fluency, F(2,57) = 0.416, p = 0.661, or number of inspirations requested between the exposure and task conditions, F(1,28) = 0.261, p = 0.612. Finally, similarly to the last study, we found a difference in inspiration influence. This time, however, the task condition displayed a higher influence than the exposure, F(1,28) = 4.59, p = 0.039 (see Table 4) .

Condition	Breadth	Depth	Influence
Baseline	4.90 (3.43)	2.40 (1.31)	-
Exposure	6.36 (2.95)	2.05 (0.785)	0.11 (0.03)
Similarity	5.78 (2.15)	2.28 (0.82)	0.13 (0.03)

Table 4. Breadth, depth, and influence for experiment 2A.

We calculated a Mixed GLM with breadth as outcome variable, condition as factor, and fluency as covariate, finding no significant difference, F(2,57) = 1.962, p = 0.150. As in the last study, we conducted a negative binomial regression for depth. With condition as factor and number of ideas as covariate, we found no significant difference, Wald Chi-Square = 2.108, p = 0.348.

Discussion for study 2A

Unlike the first study, the task condition yielded a significantly higher influence than the exposure condition, but this did not translate into an improvement in ideation breadth or depth. In general, all three conditions appeared to be very similar with respect to breadth and depth, despite the small but significant difference in influence. This is not so surprising when you consider the low number of inspirations requested for both inspiration conditions—close to 2. It is likely that the great number of ideas per inspiration either overwhelmed users or provided them with what they judged to be enough inspiration for a long stretch of time.

Experiment 2B: Smaller Inspirations, Controlled Pool

In experiment 2A, we found no meaningful difference across conditions, likely due to the very small number of inspirations requested in both experimental conditions. Therefore, we reduced the number of ideas per inspiration to 3. We also controlled the pool of ideas. One of the authors went through the existing idea pool and generated 40 different groups of 3 ideas. The goal was to create sets of ideas that shared similar elements, making the choice task easier, while at the same time having different features. For example, the idea "*Notifications such as sound or vibration when a new hit is available*" was grouped with "*sounds effects so people know when to do the surveys and also tools to see how well they are doing*" and "*The app would have alarms, bells, or sounds to notify of particular work or requesters*". The three ideas in the group would always come together, but in random order.

The method for this experiment changed in one significant way: we increased ideation time from 18 to 25 minutes, to collect more data on how ideation changes as fluency increases. We also increased the target number of users per condition (see Table 5). 89 workers participated in this study (at least 1000 completed HITs, approval > 98%, US only. In total, 863 ideas were generated. Workers also filled out a small survey at the end. Each worker was paid $3.50.

Condition	Workers	Ideas / Worker	Insp. / Worker
Baseline	35	8.43 (4.74)	-
Exposure	27	9.53 (5.95)	12.59 (12.29)
Similarity	27	9.15 (5.11)	5.74 (5.18)

Table 5. Fluency and inspiration metrics for experiment 2B.

Tables 5 and 6 summarize the metrics for this experiment. A one-way ANOVA shows a significant difference in fluency, F(2,86) = 3.528, p = 0.034. A post hoc Tukey test shows a significant difference between baseline and exposure conditions, p = 0.031, but no difference between baseline and task (p = 0.854) or exposure and task (p = 0.139). There was also a significant difference in number of inspirations between the exposure and similarity conditions, F(1,52) = 7.119, p = 0.01. However, this time no significant differences were found in inspiration influence, F(1,47) 2.019, p = 0.162.

Condition	Breadth	Depth	Influence
Baseline	5.31 (2.99)	3.17 (2.62)	-
Exposure	8.07 (3.79)	3.33 (1.68)	0.14 (0.04)
Similarity	6.33 (3.75)	2.78 (1.45)	0.16 (0.04)

Table 6. Breadth, depth, and influence for experiment 2B.

We calculated a Mixed GLM with breadth as outcome, condition as factor, and fluency as covariate, including the interaction between condition and fluency. We found a marginally significant interaction between condition and fluency, F(2,83) = 2.88, p = 0.062, but no main effect of condition on breadth, F(2,83) = 1.269, p = 0.286.

For depth, a negative binomial regression with condition as factor, fluency as covariate, and including the interaction between fluency and condition found a significant interaction between fluency and condition, Wald Chi-Square = 10.003, p = 0.007, but no significant main effect of condition, Wald Chi-Square = 4.550, p = 0.103. A pairwise comparison shows a difference only for high fluency ideators (1 SD above the mean). In this case, those in the control condition (M=6.31, SE=0.850) performed significantly above both exposure (M=3.84, SE=0.425, p = 0.009) and similarity (M=3.35, SE=0.584, p = 0.004) conditions.

We also divided both halves of the ideation and analyzed their breadth and depth separately. This was done since the effect of inspirations on users is likely not constant across the session, as they will likely be able to generate more ideas by themselves at the beginning of the session than at the end, when inspirations may be more useful. Thus, looking at the metrics over the entire session may wash out some effects.

A Mixed GLM with breadth for the first and second halves (run separately) as outcome variables, with condition as factor, and fluency as covariate yielded no main effect of condition on the first half breadth, F(2,85) =2.704, p = 0.073. On the second half, however, it yielded a main condition effect, F(2,83) = 3.527, p = 0.034, as well as a significant interaction on condition and fluency, F(2,83) = 6.957, p = 0.03. In pairwise comparisons, a difference was seen for low fluency ideators (1 SD below the mean), where the control condition (M=1.54, SE=0.322) was significantly superior to the task condition (M=0.38, SE=0.381, p = 0.022), but was not significantly different than the exposure condition (M=1.06, SE=0.447, p=0.386).

For the first half depth metric, a negative binomial regression with condition as factor, fluency as covariate, and including the interaction between condition and fluency yielded a significant main effect, Wald Chi-Square = 6.48, p = 0.039, and a significant interaction between condition and number of ideas, Wald Chi-Square = 7.46, p = 0.024. For low fluency ideators (1 SD below the mean), we see the exposure condition (M=1.99, SE=0.348) significantly outperform the control condition (M=1.05, SE=0.22, p = 0.042), but it was not significantly different to the task condition (M=1.74, SE=0.409, p=0.638). No pairwise differences were seen for high fluency ideators. The second half presented no significant interaction or main condition effect, Wald Chi-Square = 3.362, p = 0.186.

Discussion for study 2B

To summarize this study, we found a significant difference in fluency only between the exposure condition over control. The exposure condition also saw more inspiration requests. We also found baseline high fluency ideators outperforming the others in overall depth, low fluency baseline outperforming task in 2nd half breadth, and low fluency exposure outperforming baseline in 1st half depth. In other words, the inspirations not only did not help, but actually hindered the depth of high fluency ideators. It is possible that the closely related nature of the inspirations promoted fixation for them, thus detracting from their second half depth. Finally, for low fluency ideators, we see exposure helping them in first half depth, but we see tasks detracting from their second half breadth.

EXPERIMENT 3: COMPARISON ACROSS TASK TYPES

We conducted a final study in order to compare the two previous task types with a new one: combination. Combination tasks involve not only convergent processes, but also a divergent one—the generation of the new, combined idea [16]. While this can happen naturally during ideation, this task explicitly forces it to happen. Therefore, we expect a positive impact of combination on breadth. We also reverted back to completely random inspiration retrieval. The method remained the same as the one employed in experiment 2B, with the difference being that there are five conditions (control, exposure, 3 task types).

150 workers participated in this study (at least 1000 completed HITs, approval > 98%, US only), but 7 workers were not included in the analysis, as they either wrote unrelated ideas (n=1), generated unrelated tags (e.g. “tags 1”, n=4), or didn’t complete the post session questionnaire (n=2). In total, 1480 ideas were generated. Workers ideated for 25 minutes, and filled out a small survey at the end of the session. Each worker was paid $3.50.

Condition	**Workers**	**Ideas / Worker**	**Insp. / Worker**
Baseline	29	11.38 (7.178)	-
Exposure	28	10.57 (6.143)	7.70 (6.92)
Rating	27	8.48 (4.136)	4.28 (4.86)
Similarity	31	11.52 (6.45)	8.77 (5.36)
Combine	28	9.57 (5.647)	3.16 (2.51)

Table 7. Fluency and inspiration metrics for experiment 3.

Tables 7 and 8 summarize the metrics for this experiment. A one-way ANOVA shows no significant difference in fluency across conditions, F(4,142) = 1.276, p = 0.283. A one-way ANOVA for number of inspirations between conditions shows a significant difference in the number of inspirations requested, F(3,110) = 8.022, p < 0.001. A post hoc Tukey test shows that the exposure and similarity conditions were significantly higher than the rating and combination conditions (p < 0.05), but not from each other. As for influence, a one-way ANOVA shows no significant difference, F(3,105) = 1.285, p = 0.283.

Condition	Breadth	Depth	Influence
Baseline	7.86 (4.086)	3.28 (2.52)	-
Exposure	8.00 (4.830)	2.61 (1.52)	0.14 (0.05)
Rating	6.48 (3.887)	2.56 (1.76)	0.15 (0.07)
Similarity	8.74 (4.289)	2.58 (1.52)	0.12 (0.04)
Combine	8.07 (4.48)	2.07 (1.15)	0.13 (0.05)

Table 8. Breadth, depth, and influence for experiment 3.

A Mixed GLM with breadth as outcome variable, condition as factor, fluency as covariate, and including the interaction between condition and fluency yielded a significant interaction of condition and number of ideas on breadth, F(4,133) = 3.736, p = 0.006, but no main effect of condition, F(4,133) = 1.823, p = 0.128. For average fluency ideators (10 ideas), a pairwise comparison shows a significant difference between the control (M=7.20, SE=0.39) and combine (M=8.38, SE=0.39) conditions, p = 0.037. There are also significant differences for high fluency ideators (1 SD above the mean, fluency = 16 ideas), in which the control condition was outperformed by all other conditions ($p_{exposure}$ = 0.018, p_{rating} = 0.010, $p_{similarity}$ = 0.046, $p_{combine}$ < 0.001), but they were not significantly different among themselves. Figure 5 depicts the regression lines for the different conditions.

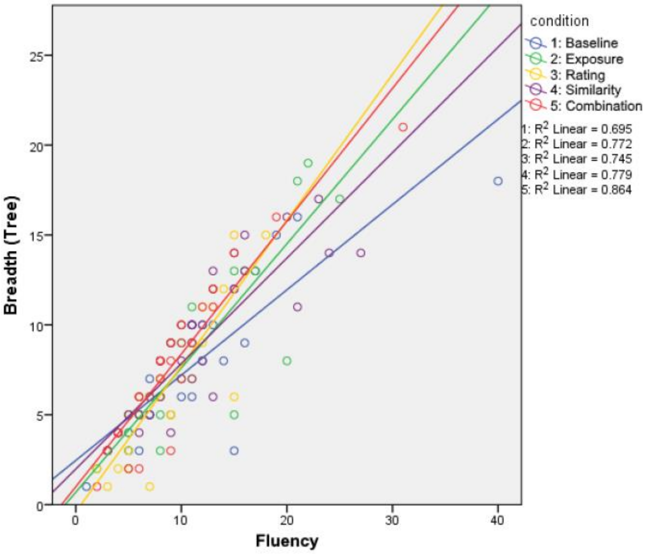


Figure 5. Regression lines for breadth by fluency.

Again, we calculated the breadth metric for each half, and found that a Mixed GLM with first half breadth as outcome variable, condition as factor, and fluency as covariate yielded no significant effect of condition, F(4,137) = 1.342, p = 0.257. However, using the second half breadth as outcome variable and including an interaction between condition and fluency yields a significant interaction between condition and fluency, F(4,133) = 7.197, p < 0.001, and a main effect of condition, F(4,133) = 2.725, p = 0.032. Figure 6 shows the marginal means for second half breadth across the different conditions, with fluency fixed at 1 SD below (4 ideas) and 1 SD above the mean (16 ideas). No significant difference is seen for low fluency ideators. For high fluency ideators, however, we see that that the three task conditions significantly outperformed the baseline (p < 0.001). When compared to the exposure condition, however, only the rating and combination conditions significantly outperformed it.

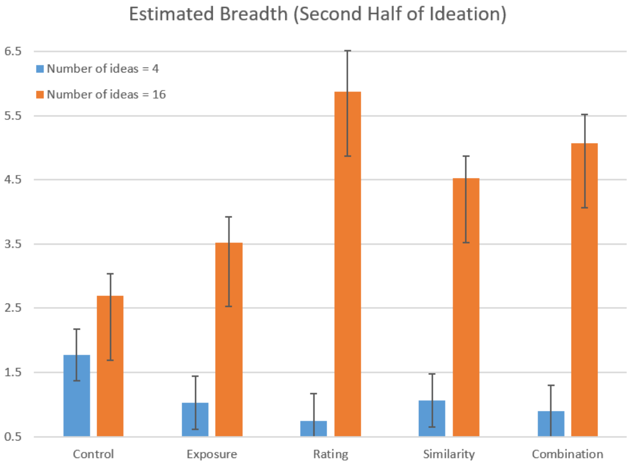


Figure 6. Marginal means and std. error for breadth (when fluency is 4 and 16) during the second half of ideation.

For overall depth, the control condition trended higher than the others, but a negative binomial regression with depth as outcome variable, condition as factor, and fluency as covariate found no effect of condition, Wald Chi-Square = 5.456, p = 0.244. The same model, but separately testing first and second half depth as outcome variables also yielded no significant differences, 1st half Wald Chi-Square = 1.469, p = 0.832, 2nd half Wald Chi-Square = 3.422, p = 0.49.

DISCUSSION

Through four experiments, we explored the integration of peripheral microtasks as part of an ideation session. Experiment 1 compared the rating task with simple exposure, finding very little influence. Experiments 2A and 2B increased the number of ideas and evaluated similarity tasks, pointing to limitations with quantity and homogeneity of inspiration ideas. Finally, experiment 3 compared all three task types together. We now discuss the main points from the combined results.

Tasks Performed as Good as Exposure, Outperforming it in Some Cases

Experiment 3 shows combination tasks outperforming the baseline for average fluency ideators. As for high fluency ideators, we find all conditions outperforming the baseline. However, when we isolate the second half of ideation, we find significant differences between the types of inspiration. The rating and combination tasks significantly outperformed the exposure condition, while similarity significantly outperformed control. One explanation for this difference is that these tasks were more cognitively demanding than the similarity and exposition inspirations. But unique characteristics of the tasks may explain them further. While usual brainstorming rules discourage criticism of ideas [25], there is evidence that criticism may foster exploration [21], which could partially account for the better performance of the rating task. Alternatively, it is possible that the rating scales provided users with a structure that guided them in generating ideas or evaluating inspirations. As for the combination task, this result was in line with our expectation, as the task also involves a divergent step [16], which could foster breadth of exploration.

Fewer Effective Inspirations May Be Better than Many Ineffective Ones

It is interesting to note that the two most effective conditions had the lowest number of inspiration requests. This may lead to the conclusion that the cognitive load of an inspiration may be more important than the number of times it is used. In other words, fewer but more effective inspirations can be better than having many less effective inspirations. An alternative explanation is that since these users requested fewer inspirations, they had more time to ideate, thus increasing breadth. However, since the fluency was not different across conditions, this is an unlikely explanation.

Inspiration Effects Depend on Timing and Fluency

The studies, especially experiment 3, highlight that inspirations may influence different users at different times. On experiment 3, for example, we see significant differences only for average or high fluency ideators. This is not surprising, as low fluency ideators may simply not be engaged enough to attend to the task or the inspirations, regardless of condition. Furthermore, results were mainly seen on the second half of ideation. This is intuitive, since at later points in time ideators are more likely to be running out of ideas [23], and thus may be more susceptible to the inspirations. This suggests that a “one size fits all” approach does not work. It may prove useful for crowd ideation support systems to restrain inspirations for a latter phase of ideation, or to initially target fluency improvement. Research that looks into adaptive support could prove fruitful.

Very Simple or Complex Inspirations Have no Effect

Studies 1 and 2A, while exploring two different types of tasks, shed light on lower and upper limits when concerning the number of ideas that can be presented for each inspiration. With both one (experiment 1) and seven (experiment 2A) ideas per inspiration, we see no significant difference between conditions. On the lower end, this lack of effect happened despite a considerable number of inspiration requests. This could be due to the simplicity of the inspirations not fostering attention to the ideas, or to users not knowing how to use the inspirations, as previously discussed. On the higher end, the lack of effect likely happened due to the low number of requests. At the end, we have found better effects with inspirations containing three ideas each. This could, however, vary depending on the inspiration type (e.g. a combination task of size 6 could be considerably more demanding than a similarity task of size 6), or even nature of ideas (homogenous idea sets may be less cognitively demanding, allowing more ideas per inspiration).

The Homogeneity of Idea Sets Can Influence the Effects

While most results were seen in breadth, we see a different pattern in experiment 2B, where the inspiration idea sets were manipulated to be more homogenous. In it, we see exposure outperforming control in first half depth, and control outperforming task in second half breadth. This could be explained partially by the homogenous nature of the ideas, as previously discussed. This indicates that the nature of the inspiration sets is highly influential in the outcome [27]. Therefore, the effect of different levels of homogeneity and task types should be explored in future work.

Some limitations of this investigation must be noted, with the first being the metrics. While the tree-based metric is consistent with previous practices and results, it needs further evaluation. A comparison with similar trees built by human experts would shed light on its performance. Alternatively, graph-based metrics could also be devised in order to better represent the inherent uncertainties in automated textual analysis (e.g. ideas could be linked to more than just one parent idea, with edge weights representing their similarity). Furthermore, we do not explore measures of creativity, whereas past research has used MTurk workers to do that [3,29,33]. However, we found workers to have very low degrees of agreement among themselves, and therefore do not report these measures. There are also limitations to the pull approach. While it allowed us to compare the performance in a natural setting, the numerical differences in inspiration requests limit our ability to clearly determine the effects of the different tasks. Finally, we do not explore the results of the tasks (e.g. the quality of the ratings). While this exploration is outside of the scope of this paper, past results are encouraging in the potential of peripheral crowd work to yield useful outcomes such as a semantic model of ideas [28].

CONCLUSION

In this paper, we have analyzed the effect of performing three different types of tasks normally done by other crowd workers: rating, similarity, and combination. We ran four subsequent experiments on MTurk to evaluate how they compare to idea exposure or individual ideation. Using breadth and depth metrics based on an ideation tree, we found the performance of task inspirations to be as good or better than simple idea exposure. We also found that the effect of inspirations depends on the fluency of ideators and the period in which it is used. Finally, we saw indications that the homogeneity of inspirations influences the outcome. Therefore, this paper provides some support and guidance in explicitly embedding microtasks into ideation, which will not only be generating information useful for convergent processes, but will also aid ideators in improving the divergence of their idea generation.

ACKNOWLEDGMENTS

The authors would like to thank all of the Mechanical Turk workers who participated in the studies. This research was funded by the CAPES Foundation, Ministry of Education of Brazil, Brasília - DF 70040-020, Brazil.

REFERENCES

1. Teresa M. Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology* 45, 2: 357.

2. Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. 2010. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, 313–322. Retrieved January 11, 2016 from http://dl.acm.org/citation.cfm?id=1866078

3. Joel Chan, Steven Dang, and Steven P. Dow. 2016. Comparing Different Sensemaking Approaches for Large-Scale Ideation. Retrieved March 11, 2016 from http://joelchan.me/files/2016-chi-sensemaking-ideation.pdf

4. Joel Chan, Steven Dang, and Steven P. Dow. 2016. Improving Crowd Innovation with Expert Facilitation.

5. Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. 2013. Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999–2008. Retrieved January 11, 2016 from http://dl.acm.org/citation.cfm?id=2466265

6. Arthur Cropley. 2006. In praise of convergent thinking. *Creativity research journal* 18, 3: 391–404.

7. Alan R. Dennis and Joseph S. Valacich. 1993. Computer brainstorms: More heads are better than one. *Journal of applied psychology* 78, 4: 531.

8. Alan R. Dennis and Mike L. Williams. 2003. Electronic Brainstorming: Theory, Research, and Future Directions. In *Group creativity: Innovation through collaboration*. Oxford University Press.

9. Michael Diehl and Wolfgang Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology* 53, 3: 497–509. https://doi.org/10.1037/0022-3514.53.3.497

10. Karen Leggett Dugosh, Paul B. Paulus, Evelyn J. Roland, and Huei-Chuan Yang. 2000. Cognitive stimulation in brainstorming. *Journal of Personality and Social Psychology* 79, 5: 722–735. https://doi.org/10.1037/0022-3514.79.5.722

11. Beth A. Hennessey and Teresa M. Amabile. 2010. Creativity. *Annual Review of Psychology* 61, 1: 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

12. Alex Ivanov and Dianne Cyr. 2006. The Concept Plot: a concept mapping visualization tool for asynchronous web-based brainstorming sessions. *Information Visualization* 5, 3: 185–191. https://doi.org/10.1057/palgrave.ivs.9500130

13. D. G. Jansson and S. M. Smith. 1991. Design Fixation. *Design Studies* 12, 1: 3–11.

14. Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 453–456. Retrieved August 17, 2015 from http://dl.acm.org/citation.cfm?id=1357127

15. Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut. 2011. Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 43–52. Retrieved August 4, 2015 from http://dl.acm.org/citation.cfm?id=2047202

16. Nicholas W. Kohn, Paul B. Paulus, and YunHee Choi. 2011. Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology* 47, 3: 554–561. https://doi.org/10.1016/j.jesp.2011.01.004

17. Aaron Kozbelt, Ronald A. Beghetto, and Mark A. Runco. 2010. Theories of creativity. *The Cambridge handbook of creativity*: 20–47.

18. Filip Krynicki. 2014. Methods and models for the quantitative analysis of crowd brainstorming. Retrieved April 5, 2016 from https://uwspace.uwaterloo.ca/handle/10012/8347

19. Richard L. Marsh, Joshua D. Landau, and Jason L. Hicks. 1996. How examples may (and may not) constrain creativity. *Memory & cognition* 24, 5: 669–680.

20. Brent A. Nelson, Jamal O. Wilson, David Rosen, and Jeannette Yen. 2009. Refined metrics for measuring ideation effectiveness. *Design Studies* 30, 6: 737–743. https://doi.org/10.1016/j.destud.2009.07.002

21. Charlan J. Nemeth, Bernard Personnaz, Marie Personnaz, and Jack A. Goncalo. 2004. The liberating role of conflict in group creativity: A study in two countries. *European Journal of Social Psychology* 34, 4: 365–374. https://doi.org/10.1002/ejsp.210

22. Bernard A. Nijstad, Michael Diehl, and Wolfgang Stroebe. 2003. Cognitive Stimulation and Interference in Idea-Generating Groups. In *Group Creativity: Innovation Through Collaboration*. Oxford University Press.

23. Bernard A. Nijstad and Wolfgang Stroebe. 2006. How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review* 10, 3: 186–213.

24. Bernard A. Nijstad, Wolfgang Stroebe, and Hein FM Lodewijkx. 2002. Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of experimental social psychology* 38, 6: 535–544.

25. Alex F. Osborn. 1963. *Applied imagination; principles and procedures of creative problem-solving*. Scribner, New York.

26. Jonathan A. Plucker and Matthew C. Makel. 2010. Assessment of Creativity. In *The Cambridge Handbook of Creativity*, James C. Kaufman and Robert J. Sternberg (eds.). Cambridge University Press, Cambridge, 48–73. Retrieved November 29, 2016 from http://ebooks.cambridge.org/ref/id/CBO9780511763205A013

27. Pao Siangliulue, Kenneth C. Arnold, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Toward Collaborative Ideation at Scale: Leveraging Ideas from Others to Generate More Creative and Diverse Ideas. 937–945. https://doi.org/10.1145/2675133.2675239

28. Pao Siangliulue, Joel Chan, Steven P. Dow, and

Krzysztof Z. Gajos. 2016. IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-Powered Real-time Semantic Modeling. 609–624. https://doi.org/10.1145/2984511.2984578

29. Pao Siangliulue, Joel Chan, Krzysztof Z. Gajos, and Steven P. Dow. 2015. Providing Timely Examples Improves the Quantity and Quality of Generated Ideas. 83–92. https://doi.org/10.1145/2757226.2757230

30. Steven M. Smith. 2003. The constraining effects of initial ideas. In *Group creativity: Innovation through collaboration*, Paul B. Paulus and Bernard A. Nijstad (eds.). Oxford University Press, New York, NY, US, 15–31.

31. Lixiu Yu, Aniket Kittur, and Robert E. Kraut. 2014. Distributed analogical idea generation: inventing with crowds. 1245–1254. https://doi.org/10.1145/2556288.2557371

32. Lixiu Yu, Aniket Kittur, and Robert E. Kraut. 2014. Searching for analogical ideas with crowds. 1225–1234. https://doi.org/10.1145/2556288.2557378

33. Lixiu Yu and Jeffrey V. Nickerson. 2011. Cooks or cobblers?: crowd creativity through combination. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1393–1402. Retrieved October 12, 2015 from http://dl.acm.org/citation.cfm?id=1979147