Providing Timely Examples Improves the Quantity and Quality of Generated Ideas

Pao Siangliulue[1]   **Joel Chan**[2]   **Krzysztof Z. Gajos**[1]   **Steven P. Dow**[2]

[1]Harvard School of Engineering and Applied Sciences
Cambridge, MA USA
{paopow, kgajos}@seas.harvard.edu

[2]Carnegie Mellon University
Pittsburgh, PA USA
{joelchuc, spdow}@cs.cmu.edu

ABSTRACT

Emerging online ideation platforms with thousands of example ideas provide an important resource for creative production. But how can ideators best use these examples to create new innovations? Recent work has suggested that not just the choice of examples, but also the timing of their delivery can impact creative outcomes. Building on existing cognitive theories of creative insight, we hypothesize that people are likely to benefit from examples when they run out of ideas. We explore two example delivery mechanisms that test this hypothesis: 1) a system that proactively provides examples when a user appears to have run out of ideas, and 2) a system that provides examples when a user explicitly requests them. Our online experiment (N=97) compared these two mechanisms against two baselines: providing no examples and automatically showing examples at a regular interval. Participants who requested examples themselves generated ideas that were rated the most novel by external evaluators. Participants who received ideas automatically when they appeared to be stuck produced the most ideas. Importantly, participants who received examples at a regular interval generated fewer ideas than participants who received no examples, suggesting that mere access to examples is not sufficient for creative inspiration. These results emphasize the importance of the timing of example delivery. Insights from this study can inform the design of collective ideation support systems that help people generate many high quality ideas.

Author Keywords

Creativity, ideation, examples, collective intelligence

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Ideation platforms—such as Quirky.com, Innocentive.com, 99designs.com—accumulate thousands of ideas contributed by their members. Because the members can see and be inspired by each other's ideas, these collections of example ideas can serve as an important resource for creative production [8]. Ideas generated by others can help innovators working on similar problems spur new concepts by broadening their notion of the design space [19, 24, 40] and allowing for reinterpretation and recombination of ideas [19, 50, 28]. When viewing ideas for inspiration, innovators should pay attention to how to select examples [24, 45, 23, 40], and how to judge their quality [19]. This is especially important because exposure to other ideas is not always inspirational: people often transfer solution elements from other ideas even when those ideas are known to be of low quality [10, 20]. Recent research shows that even experts are susceptible to such negative effects of exposure to other ideas [26]. Other ideas can also restrict one's understanding of the solution space, for example, by limiting one's ability to see novel uses for artifacts [18, 27].

Consequently, much research attention has been devoted to understanding which properties of examples are associated with inspirational outcomes. For example, research has considered how the semantic relevance [8, 9, 13], novelty [9, 1], and diversity [15, 51, 5, 45] of examples influence ideation. However, one important question has received less attention: *when* should innovators look at examples?

A variety of theoretical perspectives suggest that the impact of examples on creative output not only depends on what examples are shown but also when those examples are delivered. Cognitive theories of creative ideation suggest that ill-timed examples can disrupt a person's train of thought [34, 33] and that people benefit most from examples when they run out of ideas [43, 36, 31]. Research on flow and interruptions also suggest that automatic example delivery can be experienced as an as interruption if not timed appropriately [3, 2, 12], thereby harming creative performance. However, the literature lacks empirical tests of these hypotheses.

In this paper, we empirically test whether people benefit more from examples when they are prepared to receive them compared to seeing those same examples delivered at fixed intervals. We conducted an online ideation experiment to test two "prepared" conditions—an *On-demand* condition, in which participants determined when to see examples, and an *On-idle* condition, in which participants were automatically presented with new examples when they had been idle for a period of time. We compared these conditions against two baselines:

a condition where no examples were provided (*None*) and a condition where the examples were provided at a regular interval (*On-interval*). The baseline conditions let us distinguish the effect of access to examples *per se* from the effect of timing of the delivery of examples.

Our results show that both prepared conditions outperform the baseline conditions, but in different ways. Participants who received examples on demand produced ideas that were deemed significantly more novel by evaluators compared to participants who did not receive any examples and to participants who received examples when idle. Meanwhile, participants who received examples automatically whenever they were idle produced a larger quantity of ideas than participants in other conditions, with no significant difference in novelty compared to ideas generated by participants in either of the baseline conditions. Finally, a follow-up content analysis of the participants' ideas showed that participants who received examples on demand used examples more (i.e., borrowed/adapted more solution elements) compared to participants who received examples when idle. These results confirm that the timing of example delivery can determine the impact of examples on creative output. From a system designer's perspective, our results suggest that, instead of giving people examples in an ad hoc way, the examples should be presented at the right moment when the user is ready to make use of those examples.

RELATED WORK

Kulkarni et al. [22] examined how the timing of examples affect creative output and concluded that early or repeated—rather than late—exposure to examples improves the creativity of generated ideas. However, Kulkarni et al. delivered examples at fixed regular intervals. This may not be optimal: intuitively, one might expect that people can be more or less "prepared" to benefit from examples at different points during the ideation process.

Several theories of example use in problem solving and creative idea generation ground the intuition that people benefit more from examples when they are primed and ready. In education, the Preparation for Future Learning perspective [42, 41] posits that learners get more out of learning resources (e.g., worked examples, lectures) if they first struggle with the concepts before being exposed to those resources. Relatedly, Kapur and colleagues have shown the value of "productive failure," a two-phase instructional strategy where students first engage in generation activities (e.g., attempting to solve problems that require knowledge of the target concepts) and then engage in consolidation/instruction, where they are exposed to the target concepts in various ways [21]. These theories of learning argue that prior problem solving can prepare learners to let go of old knowledge, and prime them to notice important features of target concepts (e.g., what problem they are trying to solve).

According to a theory of idea generation called Search for Ideas in Associative Memory (SIAM) and the subsequent empirical results, example ideas can have both positive effects (cognitive stimulation) and negative effects (cognitive interference) based on when an example is shown [34, 33]. SIAM

assumes two memory systems: long-term memory (permanent with unlimited capacity) and working memory (transient with limited capacity). Long term memory is partitioned into images, which are knowledge structures composed of a core concept and its features. For example, an image can have a core concept "hotel" with features like "has rooms", "has a swimming pool", and "is cheap". When generating ideas, people run a repeated two-stage search process. First, images from long term memory are retrieved and temporarily stored in working memory. Then, in the second stage, the features of the image are used to generate ideas by combining knowledge, forming new associations, or applying them to a new domain. Retrieval of images probabilistically depends on search cues (e.g., features that are active in working memory, previously generated ideas, one's understanding of the problem). An image that is already in working memory is likely to be sampled again. SIAM, therefore, implies that seeing example ideas generally helps activate new images that would not have been accessible otherwise and thus leads to production of novel ideas. However, ill-timed examples can prematurely terminate a person's train of thought, interrupt their thinking, and cause a loss of potentially creative ideas that usually come later in the session [34, 35, 3, 2].

The Prepared Mind theory of insight offers additional insights into the optimal timing of example idea presentation. It posits that people can be more or less "prepared" to assimilate problem-relevant stimuli from the environment depending on their cognitive state [43, 36]. The theory predicts specifically that, when problem solving reaches an impasse, people maintain an open goal in memory to solve the problem, and are more motivated and better able to map problem-relevant stimuli that might have been previously ignored (e.g., because it was too semantically distant or difficult to understand/transfer). Indeed, Tseng, et al. [46] showed that people benefit more from analogically distant examples (a type of example hypothesized to be beneficial for creative inspiration [13]) during a break from problem solving after working on the problem for a significant amount of time compared to seeing the examples before working on the problem. Similarly, Moss, et al. [31] showed that people benefited more from hints after leaving a problem in an unsolved state compared to seeing the hints before working on the problem.

The shared intuition behind all of these theories is that optimal timing of example use for creative inspiration should strike a balance between allowing the ideator to queue up their own knowledge and constraints and avoiding cognitive fixation on a certain part of solution space. Therefore we predict that delivering examples *when* people run out of ideas could maximize the inspirational benefit of examples. At that point, the examples can act as external stimuli to activate new knowledge in memory to combine into new ideas. In the next section, we discuss how we might accomplish this timing of examples in an idea generation platform.

TIMING OF EXAMPLE DELIVERY

We explore two mechanisms for delivering examples to innovators when they are prepared to receive them. The first mechanism is to provide examples when people explicitly re-
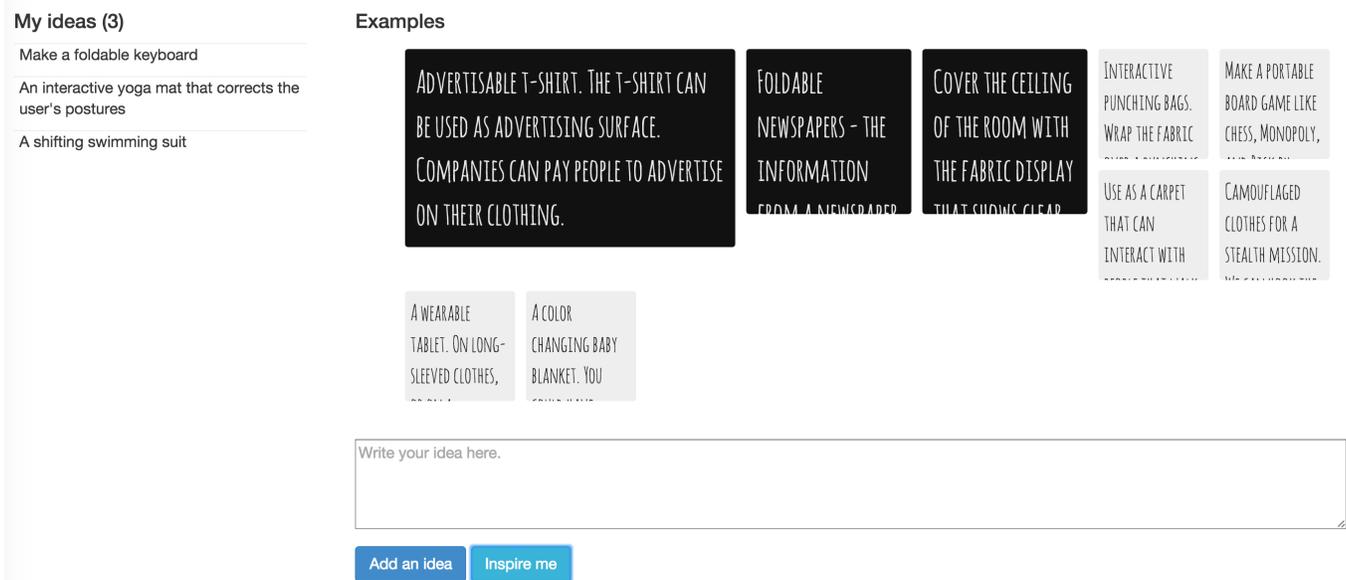
Figure 1. Screenshot of the ideation interface. Participants typed their ideas in the text box. After they submitted an idea, it appeared on the pane on the left. For those in the *On-demand*, *On-idle* and *On-interval* condition, examples were shown in the example grid above the idea entry box. The most recently received examples were shown in a big black box while earlier examples were in a gray box. The "Inspire me" button at the bottom was visible only to participants in the *On-demand* condition.

quest them (the *On-demand* condition). This approach guarantees that the examples will be provided when people are receptive to new ideas [17, 43, 36, 31]. However, people might choose suboptimal strategies for requesting examples (e.g., spending too much time looking at inspiration). People might also not be aware that they are stuck in (or biased by) old patterns of thinking [29, 7, 47] and consequently fail to request examples at an opportune time.

The second mechanism automatically provides the examples when people appear to be stuck (the *On-idle* condition). In this paper, we used a simple timeout mechanism: when no activity was detected in the interface for a fixed period of time, the system automatically provided a new set of examples of ideas generated by others. Prior research provides little guidance on how idle time during ideation relates to being in a "stuck" state. Therefore, we conducted a pilot study where we observed three people generating ideas in person. We looked at big time gaps between bursts of successive idea generation. Interviews with participants revealed that during these time gaps, they ran out of ideas on one thread and then started a new train of thought. We observed that these gaps tended to be approximately 30 seconds long. Thus, we decided on a fixed idle interval of 30 seconds for the *On-idle* condition. Analyses of time gaps before example requests in the *On-demand* condition of our main experiment provide further support for this choice of idle interval.

EXPERIMENT

Participants
We recruited 120 participants from Amazon Mechanical Turk[1] (MTurk), an online micro-labor market. Three partic-

---

[1]http://www.mturk.com

ipants did not complete the experiment and were excluded from our analysis.

We limited recruitment to workers who resided in the U.S. and who had completed at least 1,000 HITs with greater than 95% approval rate (to reduce noise from less skilled or motivated workers). Participants were paid $2.50 for their participation.

Task and Procedure
Each participant completed two idea generation tasks. In the first task, they had 3 minutes to generate as many alternative uses for rubber bands as possible. This was a warm-up task designed to familiarize participants with the system and with the example delivery mechanism. We did not include the data from this task in our analysis. In the second task, participants had 15 minutes to generate product ideas for an imaginary technology—a touch-sensitive "fabric display" that could render high resolution images and videos on any fabric through a penny-sized connector. We selected this task because it did not require extensive expertise to generate ideas, but yet was more similar to realistic design tasks than toy problems (e.g., alternative uses for a rubber band).

At the beginning of the experiment, each participant was randomly assigned to one of the four conditions:

- *On-demand*: Participants could request a new set of three examples whenever they wanted until they saw all available examples.

- *On-idle*: Participants were automatically presented with a new set of three examples when they stopped typing for 30 seconds.

- *On-interval*: Participants saw a new set of three examples at the beginning of the task and on regular intervals afterward (every minute for the alternative uses task and every three minutes for the product ideas task).

- *None*: Participants saw no examples while generating ideas.

When new examples appeared, they appeared in a set of three and were shown prominently at the top of the example grid until another set of examples came. Older examples were available throughout the idea generation session, but they were less visually prominent (Figure 1). Before each idea generation session, all participants were informed about how and when they would have an access to a new set of examples. After finishing the second task, participants filled out a survey on their demographic information and their experience during the last idea generation session.

Examples
There were 9 examples available for the alternative uses task and 15 examples for the product ideas task. Examples for the alternative uses task were obtained through an Internet search. Examples for the product ideas task were obtained from a pilot round of idea generation with 12 MTurk workers generating ideas for 15 minutes each. We selected examples as follows. A trained coder (an author) evaluated the 71 potential examples for the alternative uses task and the 60 ideas collected in a pilot study of the product idea tasks. The product ideas were coded with thematic tags like "advertising" and "camouflage." We also assessed the overall quality of each idea (judging both novelty and value). We assembled sets of three ideas that comprised both high quality and diverse theme, as both example quality and diversity have been shown to improve ideation performance [37, 25, 34, 45].

Dependent Measures And Analysis
We conducted a between-subjects study with timing of example delivery (*None*, *On-demand*, *On-idle* and *On-interval*) as the sole factor.

We collected three performance measures:

- *Number of nonredundant generated ideas.* Six redundant ideas were removed by the first author. A sample (249 raw ideas by 29 participants) was also evaluated for redundancy by the second author, and the reliability was high, ICC(2,2) = 0.83.

- *Novelty* of ideas as assessed by other MTurk workers (who were not participants in the ideation study). Previous work has also used MTurk workers to evaluate creativity of ideas (e.g., [50]).

- *Value* of ideas as assessed by other MTurk workers. This measure maps onto the dimensions of appropriateness (quality) and feasibility typically used in prior studies of creativity.

To evaluate Novelty and Value, each MTurk judge rated a random sample of 25–30 ideas. The evaluators were asked to read all ideas before rating them on 2 criteria, novelty and value, each on a 7-point likert scale. For novelty, we asked them to "consider how novel, original or surprising the idea is" (1–Not novel; 7–Very novel). For value, we asked them to "consider how useful the product idea is and how practical the idea sounds assuming the 'fabric display' technology is real" (1–Not valuable; 7—Very valuable).

Each of our evaluators rated a different subset of artifacts so calculating the agreement between evaluators is not feasible. However, we have evidence from a prior reliability study that this rating approach yields satisfactory reliability. Using three different types of creative artifacts, we measured how reliability improved as we increased the number of MTurk workers assessing creativity of any one idea. We found that a panel of three raters achieved inter-panel intraclass correlation coefficient (ICC) of 0.432. Most (98.6%) of our ideas in this study were evaluated by at least three evaluators.

To address potential misalignments in absolute means and variances in scores between evaluators, we first normalized each evaluator's scores into z-scores. We then averaged the normalized (z-)scores for each idea across evaluators. A 0 z-score meant that an idea was rated average, negative z-score means that the idea was rated below average on that criterion.

To illustrate, here are examples of ideas with low novelty (z-)scores:

- *"material for a hat"* (-1.88)

- *"games"* (-1.87)

While these are ideas with high novelty scores:

- *"Curtains that make it look like people are home when they are way. as part of a security system"* (1.78)

- *"Neckties - If they spill something on it at lunch, they can change the color so it blends in and don't have to worry about anyone noticing the stain."* (1.28)

Here are examples of ideas with low value scores:

- *"A wearable table. On long sleeved clothes."* (-1.83)

- *"A color changing bra that displays your favorite apps."* (-1.60)

While these are ideas with high value scores:

- *"Use as a stealth device for soldiers to get behind enemy lines."* (1.73)

- *"Provide to underfunded schools to replace their expensive projectors in classrooms."* (1.44)

Once they finished generating ideas, participants in the *On-demand* condition answered survey questions about when and why they requested examples (Table 1).

We also recorded timestamps when ideas got submitted and when participants saw a new set of examples. Using these timestamps, we looked at how much time passed after the latest idea submission before participants requested new examples.

Adjustments to the Data
There were originally 25 participants in the *None* condition, 26 participants in the *On-demand* condition, 31 participants in the *On-idle* condition and 35 participants in the *On-interval* condition. Our random assignment mechanism did not ensure balanced numbers across conditions because some MTurk workers abandoned the tasks when the conditions were already assigned, hindering accurate counting of participants in different conditions.

We filtered out the participants who either never requested examples or requested examples only once because these participants might not have understood that they could request examples or keep requesting examples more than once. This excluded 7 out of 26 participants from the *On-demand* condition. Because evaluating ideas is costly and the numbers of participants were unbalanced, we further randomly subsampled participants in the *On-idle* and the *On-interval* conditions so that similar numbers of participants from each condition would be used in the final analysis.

We ended up with 97 participants: 25 in the *None* condition, 19 participants in the *On-demand* condition, 28 participants in the *On-idle* condition, and 25 in the *On-interval* condition. These participants (along with their 1,149 ideas) constitute the final sample for our analysis.

RESULTS
Providing examples at idle time led to more ideas
We observed a significant main effect of timing of example delivery on the number of ideas generated by participants (F(3,93)=3.26, $p = 0.0249$). On Average, participants in the *On-idle* condition generated the most ideas (M=13.8), followed by participants in the *On-demand* condition (M=10.94), the *None* condition (M=10.88) and the *On-interval* condition (M=8.80) (Figure 2). The pairwise Student's T comparisons show significant difference between participants in the *On-idle* condition and the *On-interval* condition. There was no difference between the other pairs.
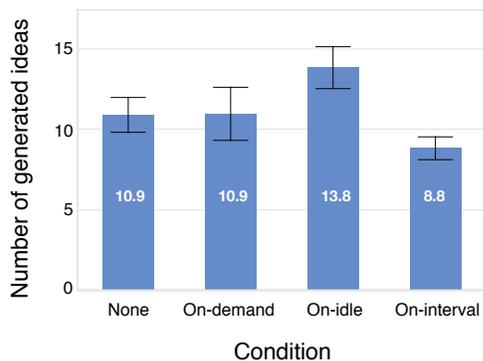

Figure 2. Participants in the *On-idle* condition generated significantly more ideas than participants in the *On-interval* condition. Error bars show standard error.

On-demand example requests led to more novel ideas
We observed a significant main effect of timing of example delivery on the average novelty of ideas (F(3,93)=4.89,

$p = 0.0034$). The pairwise Student's T comparisons show that participants in the *On-demand* condition (M=0.18) generated ideas that were deemed significantly more novel than those in the *None* condition (M=-0.18) and those in the *On-idle* condition (M=-0.01). The difference between the *On-demand* condition and the *On-interval* condition (M=0.05) was not significant (Figure 3).

We did not observe any statistically significant differences across conditions for the average value rating of ideas (F(3,93)=1.18, $p = 0.32$).
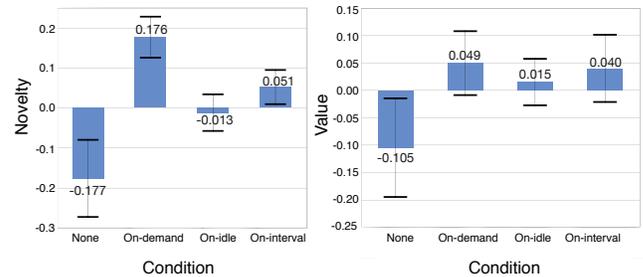

Figure 3. The mean novelty z-score for participants in the *On-demand* condition is significantly higher than for those in the *None* and *On-idle* condition. There is no statistically significant difference across conditions for the value scores. Error bars show standard error.

FOLLOW-UP ANALYSES
We conducted two sets of follow-up analyses to address questions raised by the main findings. These analyses focused on understanding why and when participants requested examples, and exploring hypotheses about why the prepared conditions (i.e., the *On-demand* and *On-idle* conditions) had different impacts on participants' creative performance.

Why and when did participants request examples?
Table 1 summarizes the survey responses of participants in the *On-demand* condition on why and when they requested examples. The responses indicate that participants primarily requested examples when they ran out of ideas. A smaller (but still sizable) proportion of participants appeared to use an alternative strategy where they looked at examples before generating ideas.

When did you request examples?	Participants N (%)
When I ran out of ideas.	15 (78.95%)
Before I started generating my own ideas.	6 (31.58%)
In the middle of coming up with new ideas.	3 (15.79%)
When I got bored.	2 (10.53%)

Table 1. When did the *On-demand* participants request examples? The majority of participants said in the post-experiment that they requested examples when they ran out of ideas.

On average, participants requested a new set of examples 31.19 seconds (SD = 44.37s) after they submitted their latest ideas (excluding example requests that came before participants submitted their first idea). This average idle time

suggests that our choice of 30s delay in the *On-idle* condition was reasonable.

However, inspecting these idle time distributions across the session yields a more nuanced picture (Figure 4). First, idle times before requesting examples tend to be shorter earlier in a session: idle times for the first and second example requests tended to be shorter than 30s. Second, there was a considerable amount of variability between participants in terms of idle times: while the mean idle time is close to 30s, participants sometimes waited more than a minute before requesting examples.
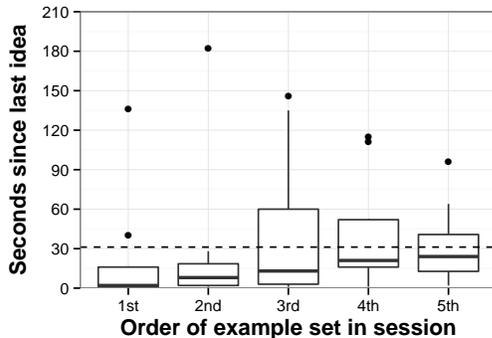
Figure 4. Boxplot of idle time before example request by order of example set in session. The mean time before requesting examples was 31.19 seconds. Participants were idle for shorter amounts of time before requesting first and second example sets than for third, fourth and fifth sets. Participants' idle times also varied considerably, with some participants waiting longer than a minute before requesting examples.

How did participants use examples?

To better understand the observed differences between the *On-demand* and *On-idle* conditions, we conducted a content analysis of the examples' impact on participants' ideas. We sampled all ideas that participants generated immediately after an example set was seen to compare against their corresponding example sets. We also included the most recent prior idea (generated within 30 seconds or less than the last seen example set) for comparison because it was common for participants to generate successive ideas within the same category or with shared functional features. In some cases, example sets were seen in succession without any ideas generated in between. In these cases, we considered the impact of the last set of examples on the next idea. This sampling procedure yielded 145 example-idea cases: 89 in the *On-idle* condition, and 56 in the *On-demand* condition. Our goal was to identify whether and how examples influenced the ideas participants generated.

The content analysis was conducted by an expert panel comprising the first and second authors. The panel separately analyzed each example-idea case to identify whether the idea appeared to be influenced by any of the examples just seen. The prior idea was included as a comparison point, since features in the idea could have plausibly been transferred/adapted from a prior idea, rather than from one or more of the examples [33]. We only considered features shared with examples that did not overlap with those of the prior idea. Specifi-

cally, we considered two kinds of example influence, following cognitive theories about example use [6, 4]:

1. Transfer of *structural features*, where the panel agreed that the idea appeared to contain mechanisms or functions (e.g., interactivity, simulation, tailoring displays to states of a system, sensing user states) also present in one or more of the preceding examples (and absent in the prior idea). For example, the idea *"Safety warnings from public institutions i.e. different colored flags on the highway that reflect Amber Alerts or how safe the roads are (a color co[d]ed system will be in place)."* shares the same mechanism of displaying the state of the systems or environment with *"Stuff animals with emotions. Make stuff animals out of this fabric. They can smile when hugged or make different facial expressions"*.

2. Transfer of *surface features*, where the panel agreed that the idea appeared to share application contexts (e.g., use for health/exercise, sports/games, learning/education) and basic features (e.g., positioning on clothing/furniture) also present in one or more preceding examples (and absent in the prior idea). For example, the idea *"To have beating organs on the outside of your clothing"* shares the same domain concept—body organ—with *"Attached with a sensor to detect body heat or heart rate, the fabric can make for clothes that detect if you are stressed out or fatigued. It will display peaceful images in soothing colors when you are stressed out"*.

Structural and surface features were considered separately to examine the possibility that participants in the *On-demand* condition were generating more novel ideas by engaging in far transfer (i.e., transferring structural features but not surface features [13]). The panel also took note of the *number* of examples that appeared to have influenced the idea.

The panel was blind to condition throughout the analysis. The panel first identified a list of features considered to be structural and surface. Then, the panel analyzed each example-idea case in an iterative manner with discussions progressing until resolution was reached. Earlier coded cases were reanalyzed in light of insights gained from later cases.

Out of the 145 example-idea cases, only 4 ideas (from 2 participants from the *on-idle* condition) were identical (or nearly identical) to the given examples. For example, a participant generated idea *"A flag that changes between various nations."* when they saw an example *"A multinational flag. Instead of having more than different flags for different nations, you can save space by having one flag that rotate showing flags of various nations."* We further inspected the ideas of these two participants and found that the copied ideas made up only a small portion of their generated ideas. The panel did not count structural or surface transfers from these copied ideas.

Figure 5 shows transfer rates for the *On-idle* and *On-demand* conditions (averaged across participants). A simple z-test for a difference in proportions yields a significant coefficient(z=3.82, $p < .001$), indicating that transfer was observed in a statistically higher proportion of cases in the *On-demand*

condition compared to the *On-idle* condition. This data suggest that participants in the *On-demand* condition used examples more often than participants in the *On-idle* condition.
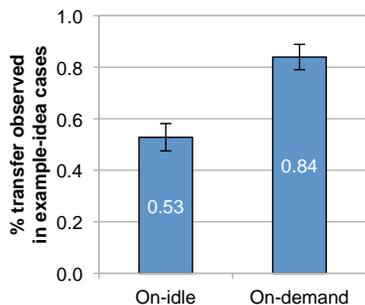
Figure 5. Participants in the *On-demand* condition used more examples to generate new ideas than those in the *On-idle* condition.

Analysis by type of feature transfer yielded similar results (see Figure 6). Transfer rates were higher for *On-demand* cases for both structural (z=2.55, $p < .05$) and surface features (z=3.68, $p < .001$). Importantly, the ratio of structural to surface transfers was similar for both conditions. These findings suggest that differences in novelty between the *On-demand* and *On-idle* conditions may be due to quantitative (i.e., more cases of examples actually influencing ideation) rather than qualitative differences (e.g., more sophisticated transfer) in how the participants used the examples.
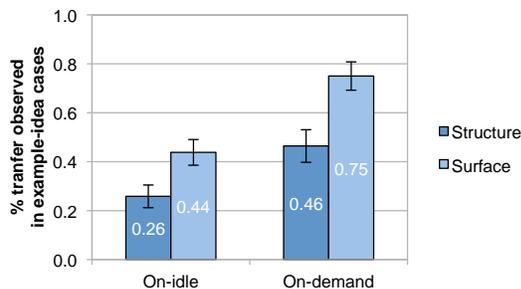
Figure 6. Participants in the *On-demand* condition transferred both structural and surface features from examples more often than those in the *On-idle* condition.

.

DISCUSSION

Adding to prior work showing the importance of considering *what* examples to see, our results demonstrate the importance of carefully considering *when* to see examples. Giving participants acces to examples on demand led to the highest ratings for novelty (but did not boost productivity). Automatically presenting examples to participants whenever they were idle also benefited ideation, but only for number (and not novelty/value) of ideas. In contrast, participants who received examples at regular intervals produced the fewest ideas (even fewer than participants who saw no examples at all). We now unpack these findings in more detail and draw out their implications for cognitive theories of creativity and the design of platforms for collaborative inspiration.

Why were on-demand and on-idle effects so different?
Why were there differences in the novelty of generated ideas between the *On-demand* and the *On-idle* conditions given that both interventions aimed to offer examples to people when they were stuck in a mental rut? One possible explanation may be related to our specific mechanism for automatically inferring when the person was stuck. Delivering examples when a person is idle for 30 seconds might be too simple or we might not have picked the right threshold time to infer the stuck moment. Our follow-up analyses of the idle timing data from the *On-demand* condition showed that the average waiting time was approximately 30 seconds, suggesting that, on average, our choice of idle interval was reasonable. Nevertheless, there was also variability in the wait times, both between participants and within sessions. While we do not believe the pattern of effects of on-idle examples is idiosyncratic to our choice of idle interval, future work exploring more nuanced idle intervals might yield more precise estimates of the size of these effects.

From a theoretical perspective, a more interesting alternative explanation might be that awareness is a key component of a prepared mind: that is, to benefit from inspirational stimuli, being stuck is not enough—you must also know that you are stuck. Theoretically, our results suggest that theories of creative insight inspiration (such as the Prepared Mind theory [43, 36, 31]) should pay more attention to metacognitive factors (e.g., awareness of one's own cognitive states). Practically, interventions designed to increase metacognitive awareness (e.g., mindfulness training) may help people maximize opportunities for inspiration. Future experiments might explore if on-idle inspiration delivery combined with such interventions could match the benefits of on-demand example delivery.

Alternatively, perhaps participants in the *On-idle* condition benefited less from examples because the examples were delivered while they were still productively accessing knowledge within a given category, even if they were not typing into the system. Our example sets were diverse and would probably have required participants to switch categories in order to recombine them into new ideas. SIAM theory [34, 33] predicts that switching categories requires effort, and can lead to productivity losses. Perhaps *On-idle* inspiration delivery could still be beneficial if the examples were "personalized" (e.g., coherent extensions of a user's current solution path). Such examples could activate other knowledge that is related to currently activated knowledge. Prior work has suggested that deep exploration within a category is an alternative (and often overlooked) pathway to highly creative ideas [32, 39]. Future work could develop novel mechanisms for real-time semantic analysis of participants' solution paths, and conduct experiments to test whether personalized inspiration could further help people benefit from inspirational examples.

Although participants in the *On-idle* condition produced ideas that were rated as slightly less novel than those generated by participants who received examples on demand, they were the most productive. This result suggests that we can prime peo-

ple to produce more ideas by showing them examples when they are idle without sacrificing the novelty or value of generated ideas. This productivity gain might be explained by the fact that new examples were presented to them before they realized that they were stuck, allowing them to pursue a new train of thought sooner instead of wasting time waiting for new ideas. However, the follow-up analysis suggested that participants in the *On-idle* condition did not use examples to guide their ideation as often as the *On-demand* participants. An alternative explanation that is more consistent with the data is that the appearance of a new set of examples signaled to people that their performance was being monitored and thus nudged them to keep on working. Prior work has shown that people increase their rate of idea generation when they know their work is being watched or will later be evaluated [48, 30, 44]. However, there is little evidence that this increased productivity also leads to higher quality (or more novel) ideas. Indeed, people often refrain from exploring "wild ideas" if they know or perceive that they are being evaluated for their ideas, a phenomenon known as evaluation apprehension [14, 11]. Future work that explores idea generation systems with automatic example delivery mechanisms should test this alternative explanation, and carefully consider participants' perceptions of automated support when designing such systems.

Did examples really help?

One important question to consider in interpreting the results is whether the examples really helped. For example, did participants in the *on-demand* condition merely copy features from the examples? Our follow-up content analysis suggests that they did indeed use examples to guide their idea generation to a greater extent than the *on-idle* participants: does this mean then that they were not being creative? One thing we can rule out is that participants were simply copying the examples wholesale. In additional follow-up analyses of ideas generated in the *on-idle* and *on-demand* conditions, participants usually generated ideas that shared features with examples instead of simply copying them. Even in rare cases when participants submitted an idea that was almost identical to the examples, subsequent ideas were their own original ideas. We suspect that submitting ideas very similar to examples helped jolt their train of thoughts.

However, ruling out wholesale copying still leaves the question of whether ideas generated by solution transfer can be considered creative. We agree with other authors [28, 38] that solution transfer *per se* does not mean that the resulting ideas are not creative (or were not produced by a creative process). Cognitive research strongly suggests that all idea generation is inevitably structured by prior knowledge [47], and studies of real-world creative behavior underscore the central importance of building on prior knowledge [16, 19]. When this structuring and solution transfer leads to ideas that are novel and valuable, we say that the idea was "inspired by" or "built upon" the example(s) [28, 19]; in contrast, when the results are less desirable, we say that the designer was "fixated" by the examples [26, 38]. Here, the fact that the *on-demand* participants mostly generated more novel ideas (and did not

merely copy examples) suggests the former interpretation of the effects of examples is appropriate.

Further insights into the potential harm of examples

Our results also join prior work in highlighting the potential negative effects of examples. Here, we add the insight that at least some of the negative effects of examples may be due to when they are seen. Although participants in the *On-interval* condition generated ideas that were no less novel than those in the *On-demand* condition, they were the least productive (even less productive than people who saw no examples at all). One potential explanation—consistent with the SIAM model—might be that the examples were experienced as interruptions or distractions, rather than inspiration; much prior work has demonstrated that interruptions are detrimental to performance [3]. Some authors have also suggested that interruptions and distractions can be especially detrimental when one is in a state of heightened focus and concentration on a creative task [12]. While this effect might be caused by our choice of time interval, this result does demonstrate that it is possible to harm productivity with ill-timed example delivery. More in-depth examination of the effect of different length of time interval could shed some light on whether negative effects of fixed interval example delivery stem from poorly selected time intervals, or whether any fixed interval example delivery is likely to be suboptimal.

CONCLUSION

In this paper we explored the question of how the impact of examples changes depending on when they are seen during ideation. We conducted an online experiment exploring two mechanisms for delivering examples at the right moment: a system that provides examples upon request and a system that proactively provides examples when a user is idle. Our results show that people benefit most from examples when they are prepared for it. Showing examples to people when they have been idle for a period of time helps people come up with more (but not necessarily better) ideas, while showing examples on-demand helps people come up with more novel ideas. In contrast, ill-timed example delivery might harm productivity, leading to fewer ideas.

These findings help support and refine theories of creative inspiration. Future work can explore different variation of time intervals for the *On-idle* and *On-demand* conditions. The length of "stuck" time interval might vary and depend on the stage of idea generation. Our findings also point toward the benefits of personalized examples that would be most helpful to people at a specific point in time. The examples in our experiment were ideas generated by participants from the pilot study or collected from the Internet search by the authors. In more realistic settings, these examples can come from various sources such as a personal collections or existing related public idea repositories where there are thousands of ideas available. Or, if people are generating ideas in groups, these inspirations can be ideas generated by others. Once a system gathers these ideas, it can select which one to show using existing methods to select a set of inspiring examples [49, 45]

We plan to explore novel mechanisms for real-time semantic analysis of people's idea exploration to gain a deeper understanding of how to best provide them with inspiration. We hope that this line of research will lay the foundation for a new generation of intelligent idea-generation support systems that augment human creativity.

ACKNOWLEDGEMENTS

REFERENCES

1. Agogué, M., Kazakçi, A., Hatchuel, A., Masson, P., Weil, B., Poirel, N., and Cassotti, M. The impact of type of examples on originality: Explaining fixation and stimulation effects. *The Journal of Creative Behavior 48*, 1 (2013), 1–12.

2. Bailey, B. P., and Iqbal, S. T. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM TOCHI 14*, 4 (2008), 21.

3. Bailey, B. P., Konstan, J. A., and Carlis, J. V. Measuring the effects of interruptions on task performance in the user interface. In *Proc. IEEE SMC*, vol. 2, IEEE (2000), 757–762.

4. Ball, L. J., Ormerod, T. C., and Morley, N. J. Spontaneous analogising in engineering design: a comparative analysis of experts and novices. *Design Studies 25*, 5 (2004), 495–508.

5. Baruah, J., and Paulus, P. B. Category assignment and relatedness in the group ideation process. *Journal of Experimental Social Psychology 47*, 6 (2011), 1070–1077.

6. Bearman, C., Ball, L. J., and Ormerod, T. C. An exploration of real-world analogical problem solving in novices. In *Prog. CogSci'02* (2002).

7. Bilalić, M., McLeod, P., and Gobet, F. Why good thoughts block better ones: The mechanism of the pernicious einstellung (set) effect. *Cognition 108*, 3 (2008), 652–661.

8. Chan, J., Dow, S., and Schunn, C. Conceptual distance matters when building on others' ideas in crowd-collaborative innovation platforms. In *Proc. CSCW'14*, ACM (2014), 141–144.

9. Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K. On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of mechanical design 133*, 8 (2011), 081004.

10. Chrysikou, E. G., and Weisberg, R. W. Following the wrong footsteps: fixation effects of pictorial examples in a design problem-solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition 31*, 5 (2005), 1134.

11. Cooper, W. H., Gallupe, R. B., Pollard, S., and Cadsby, J. Some liberating effects of anonymous electronic brainstorming. *Small Group Research 29*, 2 (1998), 147–178.

12. Csikszentmihalyi, M. *Flow and the Psychology of Discovery and Invention*. HarperPerennial, New York, NY, 1997.

13. Dahl, D. W., and Moreau, P. The influence and value of analogical thinking during new product ideation. *Journal of Marketing Research 39*, 1 (2002), 47–60.

14. Diehl, M., and Stroebe, W. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology 53*, 3 (9 1987), 497–509.

15. Doboli, A., Umbarkar, A., Subramanian, V., and Doboli, S. Two experimental studies on creative concept combinations in modular design of electronic embedded systems. *Design Studies 35*, 1 (1 2014), 80–109.

16. Eckert, C., and Stacey, M. Fortune favours only the prepared mind: Why sources of inspiration are essential for continuing creativity. *Creativity and Innovation Management 7*, 1 (1998), 9–16.

17. Friedman, R. S., Fishbach, A., Förster, J., and Werth, L. Attentional priming effects on creativity. *Creativity Research Journal 15*, 2-3 (2003), 277–286.

18. German, T. P., and Barrett, H. C. Functional fixedness in a technologically sparse culture. *Psychological Science 16*, 1 (2005), 1–5.

19. Herring, S. R., Chang, C.-C., Krantzler, J., and Bailey, B. P. Getting inspired!: understanding how and why examples are used in creative design practice. In *Proc. CHI'09*, ACM (2009), 87–96.

20. Jansson, D. G., and Smith, S. M. Design fixation. *Design studies 12*, 1 (1991), 3–11.

21. Kapur, M. Productive failure. *Cognition and Instruction 26*, 3 (2008), 379–424.

22. Kulkarni, C., Dow, S. P., and Klemmer, S. R. Early and repeated exposure to examples improves creative work. In *Design Thinking Research*. Springer, 2014, 49–62.

23. Kumar, R., Satyanarayan, A., Torres, C., Lim, M., Ahmad, S., Klemmer, S. R., and Talton, J. O. Webzeitgeist: Design mining the web. In *Proc. CHI'13*, ACM (2013), 3083–3092.

24. Lee, B., Srivastava, S., Kumar, R., Brafman, R., and Klemmer, S. R. Designing with interactive example galleries. In *Proc. CHI'10*, ACM (2010), 2257–2266.

25. Leggett Dugosh, K., and Paulus, P. B. Cognitive and social comparison processes in brainstorming. *Journal of Experimental Social Psychology 41*, 3 (May 2005), 313–320.

26. Linsey, J., Tseng, I., Fu, K., Cagan, J., Wood, K., and Schunn, C. A study of design fixation, its mitigation and perception in engineering design faculty. *Journal of Mechanical Design 132*, 4 (2010), 041003.

27. Maier, N. R. Reasoning in humans. ii. the solution of a problem and its appearance in consciousness. *Journal of comparative Psychology 12*, 2 (1931), 181–194.

28. Marsh, R. L., Landau, J. D., and Hicks, J. L. How examples may (and may not) constrain creativity. *Memory & Cognition 24*, 5 (1996), 669–680.

29. Marsh, R. L., Landau, J. D., and Hicks, J. L. Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition 23*, 4 (1997), 886.

30. Michinov, N., and Primois, C. Improving productivity and creativity in online groups through social comparison process: New evidence for asynchronous electronic brainstorming. *Computers in Human Behavior 21*, 1 (1 2005), 11–28.

31. Moss, J., Kotovsky, K., and Cagan, J. The influence of open goals on the acquisition of problem-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition 33*, 5 (2007), 876.

32. Nijstad, B. A., De Dreu, C. K., Rietzschel, E. F., and Baas, M. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology 21*, 1 (2010), 34–77.

33. Nijstad, B. A., and Stroebe, W. How the group affects the mind: A cognitive model of idea generation in groups. *Personality and social psychology review 10*, 3 (2006), 186–213.

34. Nijstad, B. A., Stroebe, W., and Lodewijkx, H. F. M. Cognitive stimulation and interference in groups: Exposure effects in an idea generation task. *Journal of Experimental Social Psychology 38*, 6 (2002), 535–544.

35. Parnes, S. Effects of extended effort in creative problem solving. *Journal of Educational psychology 52*, 3 (1961), 117–122.

36. Patalano, A. L., and Seifert, C. M. Memory for impasses during problem solving. *Memory & Cognition 22*, 2 (1994), 234–242.

37. Paulus, P. B., and Dzindolet, M. T. Social influence processes in group brainstorming. *Journal of Personality and Social Psychology 64*, 4 (1993), 575.

38. Purcell, A. T., and Gero, J. S. Design and other types of fixation. *Design Studies 17*, 4 (1996), 363–383.

39. Rietzschel, E. F., Nijstad, B. A., and Stroebe, W. Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of Experimental Social Psychology 43*, 6 (11 2007), 933–946.

40. Ritchie, D., Kejriwal, A. A., and Klemmer, S. R. d. tour: Style-based exploration of design example galleries. In *Proc. UIST'11*, ACM (2011), 165–174.

41. Schwartz, D. L., Chase, C. C., Oppezzo, M. A., and Chin, D. B. Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology 103*, 4 (2011), 759.

42. Schwartz, D. L., and Martin, T. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction 22*, 2 (2004), 129–184.

43. Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., and Yaniv, I. Demystification of cognitive insight-opportunistic assimilation and the prepared-mind perspective. In *The Nature of Insight*, R. J. Sternberg and J. E. Davidson, Eds., vol. 124. MIT Press, Cambridge, MA, 1995, 65–124.

44. Shepherd, M. M., Briggs, R. O., Reinig, B. A., Yen, J., and Nunamaker Jr, J. F. Invoking social comparison to improve electronic brainstorming: Beyond anonymity. *Journal of Management Information Systems 12*, 3 (1995), 155–170.

45. Siangliulue, P., Arnold, K. C., Gajos, K. Z., and Dow, S. P. Toward collaborative ideation at scale—leveraging ideas from others to generate more creative and diverse ideas. In *Proc. CSCW'15* (2015).

46. Tseng, I., Moss, J., Cagan, J., and Kotovsky, K. The role of timing and analogical similarity in the stimulation of idea generation in design. *Design Studies 29*, 3 (2008), 203–221.

47. Ward, T. B. Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology 27*, 1 (1994), 1–40.

48. Weber, B., and Hertel, G. Motivation gains of inferior group members: a meta-analytical review. *Journal of personality and social psychology 93*, 6 (2007), 973.

49. Xu, A., and Bailey, B. A reference-based scoring model for increasing the findability of promising ideas in innovation pipelines. In *Proc. CSCW'12*, ACM (2012), 1183–1186.

50. Yu, L., and Nickerson, J. Cooks or cobblers?: crowd creativity through combination. *Proc. CHI'11* (2011).

51. Zeng, L., Proctor, R. W., and Salvendy, G. Fostering creativity in product and service development: validation in the domain of information technology. *Hum Factors 53*, 3 (6 2011), 245–70.