Semantically Far Inspirations Considered Harmful? Accounting for Cognitive States in Collaborative Ideation

Joel Chan[1], Kanya "Pao" Siangliulue[2], Denisa Qori McDonald[3], Ruixue Liu[3], Reza Moradinezhad[3], Safa Aman[3], Erin Solovey[3], Kzryzstof Gajos[2], Steven Dow[4],

[1]Carnegie Mellon University, [2]Harvard University, [3]Drexel University,
[4]University of California San Diego,
joelchuc@cs.cmu.edu, {paopow, kgajos}@seas.harvard.edu,
{denisa.qori, rl498, reza.mora, sra68, erin.solovey}@drexel.edu, spdow@ucsd.edu

ABSTRACT
Collaborative ideation systems can help people generate more creative ideas by exposing them to ideas different from their own. However, there are competing theoretical views on whether and when such exposure is helpful. Associationist theory suggests that exposing ideators to ideas that are semantically far from their own maximizes novel combinations of ideas. In contrast, SIAM theory cautions that systems should offer far ideas only when ideators reach an impasse (a cognitive state in which they have exhausted ideas within a particular category), and offer near ideas during productive ideation (a cognitive state in which they are actively exploring ideas within a category), which maximizes exploration within categories. Our research compares these theoretical recommendations. In an online experiment, 245 participants generated ideas for a themed wedding; we detected and validated participants' cognitive states using a combination of behavioral and neuroimaging data. Receiving far ideas during productive ideation resulted in slower ideation and less within-category exploration, without significant benefits for novelty, compared to receiving no inspirations. Participants were also more likely to hit an impasse when receiving far ideas during productive ideation. These findings suggest that far inspirational ideas can harm creativity if received during productive ideation.

Author Keywords
Creativity; creativity support tools; brainstorming; examples; collaborative ideation

C&C '17, June 27 - 30, 2017, Singapore, Singapore

DOI: http://dx.doi.org/10.1145/3059454.3059455

ACM Classification Keywords
Human-centered computing~Collaborative and social computing

INTRODUCTION
Large-scale collaborative ideation platforms, like Climate CoLab and OpenIDEO, draw hundreds to thousands of contributors to collaboratively generate and develop solutions for creative problems. The promise of these platforms is that more breakthrough ideas can be developed by facilitating collaboration and remixing ideas at much higher levels of scale and diversity than before.

However, the scale and diversity of crowd ideation also presents unique challenges for collaborative inspiration. In small groups, ideation can be improved by simply exposing all contributors to all ideas [14]; however, at crowd scale, it is not uncommon to have hundreds or thousands of contributions. Ideators do not have sufficient time or cognitive resources to sift through that many ideas to select and build on ideas that are most helpful for their thinking [26,27]. Instead, ideators in these settings often resort to superficial processing of a few ideas [26,27]. Consequently, systems that can find and deliver potentially inspiring content to ideators are an important area of technical research and development for large-scale collaborative ideation platforms.

From a technical standpoint, significant progress has been made on the problem of how to structure large collections of ideas to enable exploration and navigation of the solution space [1,18,23,33,43,44]. Other studies have explored how dedicated community managers [6] or facilitators [8] might deliver appropriate inspiration to ideators.

In this paper, we consider a central *human factors* question facing designers of these inspiration delivery systems: how should inspiration delivery systems take into account the *semantic distance* of other people's ideas from the target user's ideas? Are the benefits of collaborative inspiration maximized by promoting cross-pollination of ideas (e.g., exposing ideators to ideas that are very different from their own), or by promoting deeper exploration of shared solution approaches (e.g., iterating on each other's ideas within a particular semantic region of the solution space)? Understanding the relative impact these approaches on the quantity, novelty, and diversity of generated ideas can inform the design of these platforms. To preview our results, we find that semantically far inspirations can negatively impact creative processes (speed, amount of within-category exploration) and products (novelty of ideas) if they are received during particular cognitive states.

Theoretical Foundations

Before we describe our experiment and results in detail, it is useful to consider the theoretical foundations of our investigation. Two prominent theories of creativity offer different answers to the question of semantic distance of inspirational stimuli from target user's own ideas. We selected these theories for relevance to the specific question of inspiration delivery. Our goal is not to arbitrate between competing *overall* theories of creativity, but rather to advance theoretical foundations for designing effective inspiration delivery systems.

Associationist Theory of Creativity

On the one hand, the ***associationist*** theory of creativity [21,31,34,42] argues that creativity arises from combining ideas that are very different from one another. From a cognitive standpoint, the associationist view notes a few key mechanisms by which this can be accomplished. First, creators can use *serendipity* to activate disparate portions of semantic memory at the same time. Mednick [34] noted an illustrative example of a physicist who "reduced serendipity to a method by placing in a fishbowl large numbers of slips of paper, each inscribed with a physical fact. He regularly devotes some time to *randomly* drawing pairs of these facts from the fishbowl, looking for new and useful combinations" (p. 221-222; *emphasis* ours). A related mechanism is *mediation*, which connects disparate concepts by finding deep structural similarities between them (e.g., by analogy [19]). It is worth mentioning that the original associationist theory also cites a mechanism of *similarity,* which connects disparate concepts by "spreading activation" through similar/related concepts in between. This mechanism is not shared by other associationist theories (e.g., conceptual combination, analogy), so we focus on the first two mechanisms—serendipity and mediation—that emphasize *directly* connecting semantically disparate concepts.

These theoretical assertions imply that, to improve creativity, inspiration should maximize the probability of making interesting remote associations, either by directly providing candidate combinations, or by activating distant portions of semantic memory through semantic priming [12,35]. This can be accomplished by delivering ideas that are semantically distant from the user's own ideas. According to the associationist theory, presenting ideators with such semantically far stimuli should increase the diversity and novelty of ideas they subsequently generate.

Search for Ideas in Associative Memory (SIAM) Model

On the other hand, the ***SIAM*** *(Search for Ideas in Associative Memory)* model of creative idea generation [37] argues that the answer depends on the user's cognitive state. The SIAM model posits that ideation proceeds by alternating between two kinds of cognitive states: *productive ideation* and *impasses.* During *productive ideation*, an ideator fluently accesses idea components from memory, and actively develops new ideas from those idea components. In this state, temporally adjacent ideas are often relatively close in semantic space, with some being close variations or elaborations of prior ideas, and some assembled from neighboring components in associative memory. In this state, ideators can develop more novel and useful ideas through deeper iteration and elaboration [11,15,36,41]. After exhausting resources within a semantic region, people enter an *impasse* state, and commence search for new semantic regions of memory from which to sample idea components. Idea generation during this state is slower and more effortful (as measured by time intervals between ideas, and subjective reports and/or neurophysiological measures of cognitive effort), and temporally adjacent ideas tend to be semantically distant from each other. Overall, SIAM posits that, in addition to exploring new semantic regions (when appropriate), creativity can be maximized with fluent exploration within categories, which enables people to move past common, shallow ideas to more interesting, less obvious ideas.

The theoretical assertions of the SIAM model imply that semantic distance should be accounted for in different ways for these states. During productive ideation, inspiration should enrich the local semantic region directly, or activate other (potentially valuable) idea components in the near neighborhood through semantic priming [12,35]. This can be accomplished by delivering conceptually near stimuli, which can help people more deeply explore the local region (beneficial for reaching more creative ideas [11,15,36,41]). In this state, exposure to far stimuli might even be harmful: attending to those stimuli could shift attention away from the current memory region (again, through processes like semantic priming), perhaps prematurely terminating productive chains of thought. Further, understanding and adapting far stimuli may require significant cognitive effort, and ideators may not be motivated to expend this cognitive effort if they are productively ideating within a category. During impasses, inspiration should suggest new semantic regions to explore, rather than refocusing ideators on the depleted semantic region. This can be accomplished by delivering semantically far stimuli, which could help accelerate the process of finding new productive regions to explore by activating more diffuse portions of memory.

From Theories to Inspiration Delivery System Designs

The associationist and SIAM theories therefore predict very different best and worst inspiration delivery approaches (see Table 1). Associationist theory suggests an **ALWAYS-FAR** inspiration approach would be best for ideation, where

	Predicted *best* approach	Predicted *worst* approach
Associationist theory	**ALWAYS-FAR**: deliver *far* stimuli, regardless of cognitive state; increases novelty and diversity of ideas by promoting remote associations *(SIAM predicts neutral effect on novelty: gains in novelty from pointers to new ideas offset by hindering exploration within categories)*	**ALWAYS-NEAR**: deliver *near* stimuli, regardless of cognitive state; decreases novelty and diversity of ideas by suppressing remote associations *(SIAM predicts neutral effect on novelty: loss of novelty from lack of pointers to new ideas offset by gains from fluent within-category exploration)*
SIAM model	**MATCH-STATE**: deliver *near* stimuli during productive ideation, and *far* stimuli during impasses; increases novelty of ideas by promoting fluent exploration within categories, and providing pointers to new areas of exploration at the appropriate time	**MISMATCH-STATE**: deliver *far* stimuli during productive ideation, and *near* stimuli during impasses; decreases novelty of ideas by hindering fluent exploration within categories, and suppressing pointers to new areas of exploration at the appropriate time

***Table 1*. Best and worst approaches for choosing semantic distance of inspirational stimuli, as predicted by the associationist and SIAM theories of creativity. Associationist theory predicts that an ALWAYS-FAR approach is best, and an ALWAYS-NEAR approach is worst; SIAM predicts that a MATCH-STATE approach is best, and a MISMATCH-STATE approach is worst. SIAM also makes competing predictions for the associationist theory's predicted best and worst approaches.**

the system strives to deliver semantically distant ideas to the user, without accounting for cognitive states. According to the associationist theory, the least helpful approach would be the **ALWAYS-NEAR** inspiration strategy, which would be predicted to *harm* ideation by always constraining the user to a semantically adjacent region.

In contrast, SIAM suggests a **MATCH-STATE** inspiration approach would be best for ideation, where the system delivers near stimuli during productive ideation, and far stimuli during impasses. SIAM further predicts that a **MISMATCH-STATE** inspiration approach will be least helpful, since it presents the user with the opposite of their theoretically predicted inspiration needs during each state (i.e., potentially distracting far stimuli during productive ideation, and constraining near stimuli during impasses).

Interestingly, SIAM also offers a competing prediction for the associationist's best and worst approaches: the **ALWAYS-FAR** and **ALWAYS-NEAR** approaches should yield similar levels of novelty. This is because SIAM predicts that, in the **ALWAYS-FAR** approach, the increase in novelty from providing pointers to new areas of exploration should be offset by losses in novelty due to hindered within-category exploration. In the **ALWAYS-NEAR** approach, SIAM predicts that decreased novelty from lack of pointers to new areas of exploration should be offset by increased novelty from fluent within-category exploration [5,36].

To advance research and development of inspiration delivery systems for collaborative ideation platforms, empirical work is needed to tease apart whether the associationist or SIAM theories (or neither) are more useful theoretical guides for how to appropriately account for semantic distance of potential inspirational stimuli.

Overview and Contributions of The Present Study

In this paper, we report the results of an empirical test of these theories by comparing each of their predicted best and worst inspiration delivery approaches against a **NO-STIMULI** baseline. In an online ideation experiment, 245 participants generated ideas for themed weddings in one of the four inspiration conditions (**ALWAYS-FAR**, **ALWAYS-NEAR**, **MATCH-STATE**, and **MISMATCH-STATE**), or in the **NO-STIMULI** baseline condition. We detected changes between productive ideation and impasse states through a simple self-report mechanism. This approach was validated by behavioral (ideation was significantly slower right before an impasse) and neuroimaging data (which showed neuroimaging markers of significantly elevated cognitive effort right before an impasse).

Consistent with SIAM predictions, **MISMATCH-STATE** and **ALWAYS-FAR** participants generated ideas at a slower rate than **NO-STIMULI** participants, and **ALWAYS-FAR** participants iterated less within categories (as measured by mean similarity between subsequent ideas) compared to **NO-STIMULI** participants. Further, participants who received near stimuli during productive ideation (**MATCH-STATE**, **ALWAYS-NEAR**) were less likely to face impasses than participants who received far stimuli during productive ideation (**ALWAYS-FAR**, **MISMATCH-STATE**). Contrary to associationist predictions, **ALWAYS-FAR** ideas were not significantly more diverse or novel than **NO-STIMULI** ideas; instead, **ALWAYS-FAR** ideas were marginally statistically significantly *less* novel than **NO-STIMULI** ideas.

This paper contributes new insights for how to best promote creative inspiration on collaborative ideation platforms. Specifically, our findings show that far inspirational ideas—though considered to be generally useful for creative inspiration—can harm creativity if received during productive ideation. Our findings also imply that the SIAM model's state-contingent view of inspiration needs is more useful as a theoretical starting point than the associationist theory of creativity for guiding the design of collaborative inspiration systems.

EXPERIMENT

Participants

We recruited 245 participants (mean age = 33.5 years, SD=10.4, 51% female) from Amazon Mechanical Turk (MTurk). All participants were located in the U.S. and had 95% approval on at least 100 MTurk tasks. Participants were paid $1.25 for their time (approximately $6/hr wage, given average completion times of 12–13 minutes).

Study Design

We used a between-subjects design. Participants were randomly assigned to one of the 5 conditions: 1) the **NO-STIMULI** baseline (N=54), 2) associationist theory's predicted best **ALWAYS-FAR** condition (N=47), 3) associationist theory's predicted worst **ALWAYS-NEAR** condition (N=48). 4) SIAM theory's predicted best **MATCH-STATE** condition (N=51), and 5) SIAM theory's predicted worst **MISMATCH-STATE** condition (N=45).

Brainstorming Task

Participants generated ideas for a themed wedding, where each idea consisted of 1) a theme, 2) a main prop to be used for guest activities, and 3) a freeform description of how the prop would be incorporated into the wedding. We chose this task structure to maximize our ability to accurately tailor conceptual similarity based on participants' current thinking (and therefore experimentally isolate our intervention) in real-time. Achieving real-time semantic tailoring of potential stimuli to unstructured participant ideas of varying length and specificity is challenging to accomplish with a high degree of accuracy. To address this concern, our brainstorming task is semi-structured: participants separately specify theme (e.g., "medieval") and prop (e.g., "silver spoons") components of their themed wedding idea. This allows us to perform fast and accurate tailoring based on those single or compound words where computational similarity measurements tend to do better. For example, models like Pennington et al's [38] Global Vectors for Word Representation (GloVe) model — which uses an unsupervised learning algorithm to learn vector representations for words from global word-to-word co-occurrence statistics within a corpus — are able to achieve between 60% and 84% accuracy on word analogy tasks. This brainstorming task also achieves a degree of ecological validity since developing ideas for themed weddings is a common real-world creative task.

Sampling Inspirations based on Conceptual Distance

We used pre-trained GloVe vectors (trained on approximately 6 billion tokens (Wikipedia 2014 and Gigaword 5 corpora, with 300 dimensions) provided by Pennington et al [38] to perform similarity matching. While other vector-space models like Latent Semantic Analysis (LSA) have a longer history in cognitive science for measuring semantic distance [32], we opted to use GloVe, a recent state-of-the-art model that typically agrees well with classic models like LSA [17], while being capable of modeling more nuanced semantics, such as simple four-term analogies (e.g., man:woman = king:queen) [38].

Our database of potential inspirational stimuli consisted of 455 themes and 655 props collected from pilot runs of this study (with 207 MTurk workers; none of these workers also participated in the main study).

We showed inspirations as sets of 3 themes and 3 props, assembled in real-time and tailored to participants' last generated idea. *Near* stimuli are sampled to be near in semantic space, but no nearer than cosine similarity of 0.5 in the GloVe vector space. We selected this threshold to avoid duplicates and very close matches that are likely already activated, and also to activate the periphery of the current location in the semantic network in order to enrich the semantic region with additional potentially active idea components. *Far* stimuli were sampled to be as far from current thinking as possible.

To ensure our sampling approach focused on conceptual distance, we also controlled for diversity of inspiration sets (the relative distance between inspirations *in a set*) because the diversity of inspirational examples has been shown to impact creative performance [4,43]. Conceptual distance and diversity tend to be positively correlated, but it is hard to generate diverse sets for near stimuli. All things being equal, ideas that are all close to a seed idea will be relatively close to each other, compared to ideas sampled from distant semantic regions. Therefore, we restricted the diversity of sets to be relatively low. We use a simple sampling algorithm to ensure low diversity of sets: For each query, we first sampled a seed inspiration (whether near or far). Then, we found two nearest neighbors of that seed inspiration (where the cosine similarity of those neighbors to the seed and each other were less than 0.5). This completed a set of 3 inspirations for the query.

The following are examples of near and far inspirations sampled by our approach for two different themes:

For "football", *Near*: [season, fun and games, fourth of July], and *Far*: [toga, hula, prom]. For "steam punk", *Near*: [album, light of love, rock], and *Far*: [minions, knight and damsel, ghostbusters].

Validating Stimuli Sampling Approach

To validate our approach, we randomly sampled 100 themes generated by participants in the study, along with the near and far sets of inspirations actually retrieved for those themes during the experiment. We then generated a new set of inspirations that was the opposite distance (either near or far) from each theme. A trained research assistant (blind to which sets were deemed near or far by the algorithm) then went through each theme and marked which of the two sets (left or right) was "nearest" to the theme. A second judge (one of the authors, also blind to the algorithm's predictions) completed judgments for a random subset of 40 of the items, and agreement between the human judges was very high, Cohen's $\kappa = 0.95$. The research assistant judged the remaining items. The model's selection of near stimuli corresponded well to the judge's selection, Cohen's κ = 0.84, validating our semantic tailoring manipulation. In terms of "absolute" distance, the near stimuli in our dataset were, on average, 4 nodes away from the participant's last idea in Wordnet's association network (e.g., WOLF-->canine-->carnivore-->feline-->CAT), compared to 9 nodes away for far stimuli.

To validate our diversity control mechanism, we sampled 855 inspiration sets actually provided to participants during our experiment, and measured their pairwise distances using GloVe. In this sample, far sets were not more diverse than near sets; in fact, there was a tendency for near sets to be more diverse (M=.39, *SE*=.00) than far sets (M=.27, *SE*=.00; t=18.67, p<.01). This suggests that our diversity control mechanism successfully removed the usual coupling between distance and diversity of inspiration sets.

Inferring Ideators' Cognitive States

Our experiment requires that we accurately infer participants' cognitive state at each moment. Automatic detection is likely to be noisy; people can be idle for different reasons, and may be productively thinking while not typing. In contrast, prior work shows that people can notice when they are stuck [45]. However, ideally we do not want to burden participants with constantly monitoring their own cognitive state when they are productive.

Therefore, we designed a partially user-driven approach to infer cognitive states. The default state is productive ideation. The participant triggers a state change to impasse by requesting a set of inspirations, with the intuition that participants in this state would naturally seek out new stimulation. The system then infers a state change back to productive ideation once the participant submits a new idea.

Validating User-Driven Inferring of Cognitive States

Since this user-driven approach is novel, we sought to validate that it succeeds at differentiating between cognitive states. We conducted a small pilot study in which participants brainstormed while wearing functional near-infrared spectroscopy (fNIRS) brain sensors. fNIRS is a neuroimaging method that detects changes in the concentration of oxygenated and deoxygenated blood in the brain, relative to a reference point (e.g., during resting state, or a fixed time interval prior to an "event"). These changes in blood oxygen concentration can be used to infer changes in brain activity in particular regions of the brain [49], similar to the blood oxygen level dependent (BOLD) measure used in fMRI.

Neuroimaging Validation. The fNIRS device, manufactured by ISS, Inc. contained six measurement channels with 3-cm source-detector distances. Participants experienced the same set of procedures as our online participants, except they generated ideas for 20 minutes (to maximize the amount of data points per participant), and participated in a post-task semi-structured interview. We preprocessed the

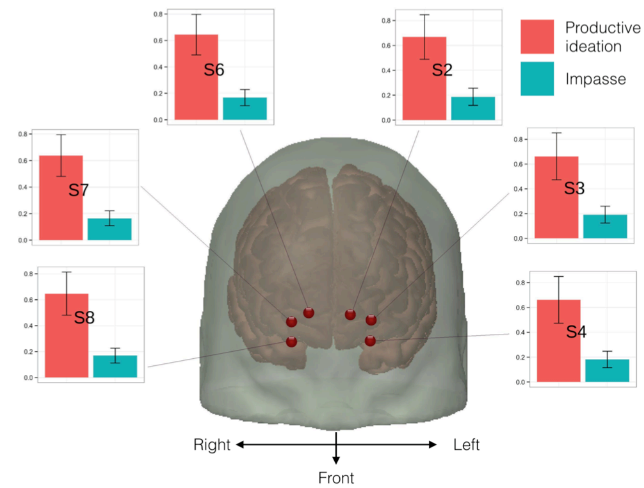

Figure 1. **Participants in our pilot study showed higher levels of cognitive effort during system-inferred impasses (green bars) compared to system-inferred productive ideation (red bars), as indicated by lower levels of ΔHbR (in micromolars, μM) in regions of the prefrontal cortex (measured using functional near-infrared spectroscopy).**

fNIRS data for analysis using Homer2 [24], a MATLAB-based application for processing fNIRS data. Preprocessing involved converting the raw light intensity data to oxygenated (HbO) and deoxygenated hemoglobin (HbR) changes. A high-pass filter was also applied at 0.5 Hz. Our sample consists of 6 participants, with a total of 23 instances of inferred impasses.

We focused our analysis on comparing brain activity 10 seconds immediately before an inspiration request (which we assume would be an impasse state) and 10 seconds immediately after the first idea submission after an inspiration request (which we assume would be a productive ideation state). Our hypothesis is that we should see brain activity that indicates higher levels of cognitive effort during the inferred impasse state, compared to the productive ideation state. The expected hemodynamic response would be a negative change in HbR during increase cognitive load [25,46]. Thus, we operationalize cognitive effort as the maximum change in HbR relative to the 2s just prior to the event.

We found a significant difference in the maximum HbR signal between the two states, with the pre-stuck period being lower than the post-stuck period across all of the channels we measured (see Figure 1; all p<.01). This finding suggests participants were exerting high levels of cognitive effort during inferred impasses (i.e., right before requesting inspirations).

Behavioral Validation. These neuroimaging results were further corroborated by comments participants made during the post-task semi-structured interview. For example, one participant said she clicked to request more inspirations *"mostly just like after I had like exhausted the ideas in my mind and I was like OK I don't know what could possibly*

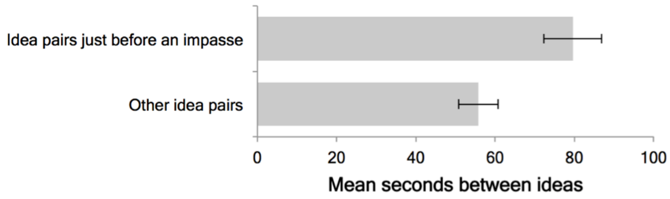

Figure 2. **Participants generated ideas more slowly just before an impasse compared to other points in their ideation session.**

be next". Another participant said he requested inspirations when he *"was just running out of ideas"*.

Further, participants in our online experiment took significantly more time between subsequent idea submissions just before an inferred impasse (M=79.6 seconds, *SE*=7.3), compared to idea transitions not temporally adjacent to an inferred impasse event (M=55.8, *SE*=5.0), t(19.4) = 3.3, *p*<.01 (see Figure 2). Note that these do not include the interval between the last idea submission before an inspiration request and the request, or the interval between the request and the first idea submission after the request. The intervals are therefore indicative of the speed with which ideas are generated, and not the time it takes to perform extra tasks, such as (deciding to) request more inspirations. This difference in inter-idea interval is consistent with SIAM's predictions of faster ideation during productive ideation and slower ideation during impasses.

Altogether, these results suggest that our user-driven approach successfully detects cognitive state changes (i.e., inspiration requests signal a transition to an impasse state).

System and Interface

Figure 3 shows the ideation interface used in our experiment. Participants enter ideas in a semi-structured format (separate fields for themes, props, and descriptions). In the left pane, the system automatically retrieves a new set of themes and props after each idea entry. The participant is assumed to be in a productive ideation state unless they click the *"Give me other inspirations"* button. When this button is clicked, the system infers the participant is in an impasse state, and retrieves a new set of inspirations accordingly. Participants can refresh the inspiration feed during the impasse state as many times as they wish. Each button click during this state retrieves a new set of inspirations. Once a participant submits a new idea after requesting inspirations, the system infers the participant has returned to a productive ideation state.

The system was built in Meteor.js (a Node.js-based web application framework). The system communicates with a similarity engine via a RESTful API to select inspirations based on semantic relatedness to participants' current idea focus. Inspiration retrievals took about 1–2 seconds.

Procedure

After providing informed consent, participants were randomly assigned to one of the 5 conditions. Next, they

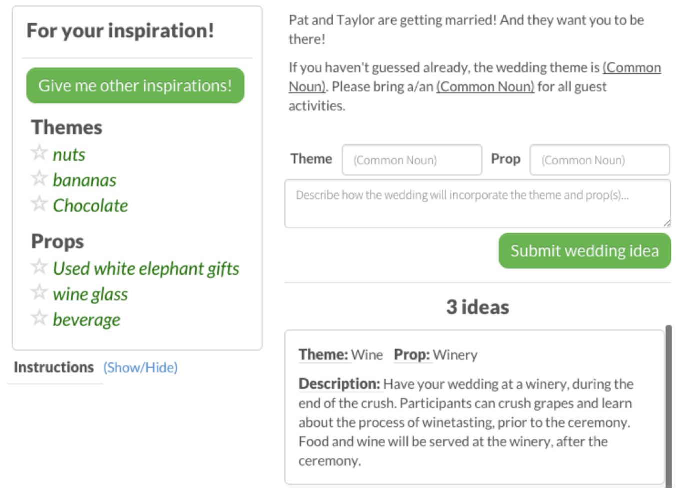

Figure 3. **Participants enter ideas in a semi-structured interface. Inspirations in the left pane automatically update during productive ideation; when participants request new inspirations, this signals an impasse state to the system, and new inspirations are retrieved. The control condition interface is identical except for the absence of an inspiration feed.**

entered a warm-up task screen where they generated alternative uses for a brick for one minute. After this, they completed a short tutorial that highlighted the key features of the interface. Finally, they generated ideas for the main problem for 8 minutes. The system automatically took them to a final survey page after 8 minutes.

MEASURES

Within-Category Fluency

We operationalize two measures of within-category fluency that capture related but distinct aspects of the theoretical construct of within-category fluency: inter-idea interval (probability of an impasse), and transition similarity.

Inter-Idea Interval

SIAM posits that ideation within a category is more rapid than generating ideas in-between categories; thus, participants who have higher within-category fluency should, on average, have shorter intervals between idea submissions. Thus, we operationalize inter-idea interval as the median number of seconds between subsequent idea submissions, as logged by our system. We report median inter-idea interval because this measure is insensitive to long inter-idea intervals during impasses and instead, reflects how rapidly a participant was generating ideas while they were in a productive ideation state. We also statistically control in our analysis for the number of ideas generated. This allows us to more cleanly capture the degree of within-category fluency.

Transition Similarity

Ideas within a category tend to be more semantically similar to each other than ideas between categories: thus, participants who have higher within-category fluency should, on average, have higher similarity between successive ideas. We operationalized transition similarity as the median GloVe cosine between themes and props of subsequent ideas. As with inter-idea interval, reporting the median allows us more cleanly capture the degree of within-category fluency.

Overall Fluency: Number of Ideas Generated

Fluency was operationalized as the number of ideas generated for the problem.

Diversity: Mean Pairwise Distance Between Ideas

Diversity was operationalized as mean pairwise distance (the reverse of similarity) between a participant's ideas as measured by GloVe.

Novelty

We recruited 185 workers from MTurk to rate the novelty of the generated ideas on a scale of 1 (Extremely Obvious) to 7 (Extremely Novel). Each worker rated a random subset of approximately 30 ideas. Computing correlations between each judges' ratings and the overall aggregate score yielded an average aggregate-judge correlation of r=.64. To deal with potential differences in usage of the rating scale across raters (e.g., some might only use the upper end of the scale), we compute standardized scores (a.k.a. "z-scores") within raters (i.e., a mean score of a rater was subtracted from that rater's each individual score and the difference was divided by the standard deviation of that rater's scores).

An example of a *high* novelty idea is "[Chemistry] [Lab experiment] The couple could conduct a common laboratory experiment combining two substances to create a third as part of their ceremony symbolizing and celebrating their union." (z-score=1.61). An example of a *low* novelty idea is "[formal] [gift] It would be a typical wedding" (z-score=–1.94)

Each participant's novelty score was the *highest* novelty score across his/her ideas. This conceptualization of novelty is a good fit for the predictions of both associationist and SIAM theories for novelty, which emphasize novelty as an outcome of the ideation process (i.e., the most novel idea that was generated, instead of average novelty of all ideas).

Control Measure: Baseline Fluency

In ideation studies, it is important to control for pre-existing differences in participants' creative capacities, such as baseline fluency of ideation [7,9]. Baseline fluency was operationalized as the number of alternative uses generated for a brick during the warm-up task that participants completed prior to the main ideation task. This measure is intended to capture both aspects of baseline creative fluency [20], and aspects of participant motivation and comfort with the interface (all important for creative productivity).

RESULTS

System-Usage Statistics

Participants generated a total of 1,574 ideas across conditions. 85% of participants in the inspiration conditions reported using the inspiration feature in some way (e.g., attending to inspirations, requesting inspirations). Of those who did, 68% self-reported interacting with the inspirations (e.g., attending to, using as inspiration) at least "somewhat frequently" (3 on a scale of 1 to 5) when they weren't actively clicking to get more inspirations, *M*=3.0 (*SE*=0.1).

Across the inspiration conditions, 49% of participants requested inspirations at least once. Interestingly, however, likelihood of an inspiration request was not equal across conditions. Participants who received near stimuli during productive ideation (**ALWAYS-NEAR** and **MATCH-STATE**) were less likely to request an inspiration at least once (M=.42, *SE*=.05) than participants who received far stimuli during productive ideation (**ALWAYS-FAR** and **MISMATCH-STATE**, M=.57, *SE*=.05). A logistic regression model, predicting the probability of inspiration request as a function of stimuli distance during productive ideation, showed that the difference between the near and far distance groups was statistically significant, z=1.96, *p*=.05. Since this analysis was conducted in response to seeing the data (vs. hypothesized in advance, as with the primary analyses in the subsequent section), we wish to clearly mark this finding as exploratory (rather than confirmatory). We revisit this finding in the Discussion.

Overall, these numbers suggest that the features of the system relating to inspiration were used reasonably frequently, providing an adequate test of our manipulations.

Primary Analyses

For each dependent measure, we estimate an ANCOVA model with baseline fluency as a control covariate (if it is a statistically significant predictor of the dependent measure). For median inter-idea interval, we also include number of ideas generated as a theoretically motivated control covariate (if it is statistically significant). All significant main effects of condition are followed up with planned contrasts against the **NO-STIMULI** condition, using Dunnett's procedure [16] to control Type I error inflation from multiple comparisons. Table 2 summarizes model-adjusted means and standard errors for each dependent measure by condition.

Slower Ideation with Always-Far and Mismatch-State

To investigate ideation pace in each condition, we estimated an ANCOVA predicting median *inter-idea interval* as a function of condition, controlling for number of ideas (which was significantly negatively correlated with inter-idea interval, r=–.63, *p*<.01). Recall that controlling for number of ideas provides a finer-grained measure of within-category fluency, allowing us to discern qualitative differences (low vs. high within-category fluency) between quantitatively similar (overall number of ideas) ideation traces. The model showed a significant main effect of condition on median seconds between subsequent ideas, F(4,233)=3.2, *p*=.01. Planned contrasts showed that **ALWAYS-FAR** and **MISMATCH-STATE** participants had significantly longer median inter-idea intervals (t=2.8, *p*=.02 and t=3.1, *p*=.01, respectively) compared to **NO-STIMULI** participants.

	Inter-idea interval	Transition similarity	Overall Fluency	Diversity	Novelty
NO-STIMULI	64.2 *(5.3)*	0.19 *(0.01)*	6.5 *(0.5)*	0.84 *(0.01)*	0.88 *(0.07)*
ALWAYS-FAR	**86.2 *(5.7)* ***	**0.12 *(0.02)* ****	6.1 *(0.5)*	0.86 *(0.01)*	**0.64 *(0.07)* ^{m}**
ALWAYS-NEAR	74.3 *(5.6)*	0.20 *(0.02)*	6.4 *(0.5)*	0.81 *(0.01)*	0.67 *(0.07)*
MATCH-STATE	76.6 *(5.5)*	0.19 *(0.01)*	7.0 *(0.5)*	0.83 *(0.01)*	0.88 *(0.07)*
MISMATCH-STATE	**88.7 *(5.8)* ****	0.14 *(0.02)*	6.5 *(0.5)*	0.84 *(0.01)*	0.79 *(0.07)*

Table 2. **Model-adjusted means and standard errors for each dependent measure by condition. ^{m} *p* < .10, * *p* < .05, and ** *p* < .01 for contrasts with the NO-STIMULI baseline, with Dunnett's correction for multiple comparisons. Contrasts show that ALWAYS-FAR resulted in significantly longer *inter-idea intervals* and significantly lower *transition similarity* and *novelty* than NO-STIMULI; MISMATCH-STATE resulted in significantly longer *inter-idea intervals* than NO-STIMULI.**

Lower Transition Similarity in Always-Far Condition

Transition similarity was not significantly correlated with baseline fluency (r=.02, *p*=.79). Therefore, we estimated an ANOVA with condition as the only factor. The model showed a significant effect of condition, F(4,218)=4.9, *p*<.01. Planned contrasts showed that **ALWAYS-FAR** participants had significantly lower median transition similarity (t=-3.1, *p*<.01) than participants in the **NO-STIMULI** condition. A post-hoc Tukey test also showed that **ALWAYS-FAR** participants had lower median transition similarity than both **MATCH-STATE** (*p*=.02) and **ALWAYS-NEAR** participants (*p*<.01).

Fluency: Equal Number of Ideas Across Conditions

An ANCOVA controlling for baseline fluency showed no significant main effect of condition on the *number of ideas*, F(4,239)=0.5, *p*=0.77.

Always-Far Leads to More Diversity than Always-Near

No theoretical covariates were statistically significantly related to *diversity*. Therefore, we estimated an ANOVA with condition as the only factor. This ANOVA showed a significant main effect of condition on diversity, F(4,233) = 2.97, *p*<.01, but planned contrasts did not show any differences between the inspiration conditions and the control condition. However, a post-hoc Tukey test showed that participants in the **ALWAYS-NEAR** conditions had significantly lower diversity of ideas compared to the **ALWAYS-FAR** condition (*p*=.01).

No Benefits for Novelty in Always-Far Condition

Baseline fluency was not significantly correlated with novelty (r=.03, *p*=.68). Thus, we estimated an ANOVA with condition as the only factor. This model showed a significant main effect of condition, F(4,239)=2.5, *p*=.04. Planned contrasts showed no significant differences between **ALWAYS-FAR** and **NO-STIMULI** participants, t=–2.3, *p*=.07; however, the mean trends were in the opposite direction predicted by the associationist theory, with the most novel ideas of **ALWAYS-FAR** participants rated as *less* novel (M=0.64, *SE*=0.07) than **NO-STIMULI** participants' most novel ideas (M=0.88, *SE*=0.07).

DISCUSSION

In this study, we explored how the semantic distance of inspirations from the target user's own ideas impacts their creative performance. Specifically, we compared two competing theoretical recommendations from creativity theories: 1) the associationist view, which predicted that always providing far stimuli would be most beneficial, and 2) the SIAM model of creative ideation, which predicted that a state-contingent inspiration delivery (where near stimuli are delivered during productive ideation, and far stimuli during impasses) would be most beneficial.

Figure 3 summarizes our findings and their implications for the two competing theories. Consistent with the associationist view, ideas generated in the **ALWAYS-FAR** condition were significantly more diverse than those generated in the **ALWAYS-NEAR** condition; however, **ALWAYS-FAR** ideas were not significantly more diverse than **NO-STIMULI** ideas. Further, contrary to associationist predictions, **ALWAYS-FAR** ideas were not significantly more novel than **NO-STIMULI** ideas; instead, the mean trends showed that **ALWAYS-FAR** ideas were possibly *less* novel than **NO-STIMULI** ideas.

In contrast, consistent with SIAM predictions, **MISMATCH-STATE** and **ALWAYS-FAR** participants generated ideas at a slower rate than **NO-STIMULI** participants, and **ALWAYS-FAR** participants iterated less within categories compared to **NO-STIMULI** participants (indicating that far stimuli hinder within-category exploration). Finally, an exploratory analysis showed that participants who received near stimuli during productive ideation (**MATCH-STATE**, **ALWAYS-NEAR**) were less likely to request inspirations (which our system used to detect impasses) than participants who received far stimuli during productive ideation (**ALWAYS-FAR**, **MISMATCH-STATE**), suggesting that near stimuli could extend productive ideation chains (relative to far stimuli). However, **MATCH-STATE** participants did not have greater within-category fluency, overall fluency, or novelty of ideas than **NO-STIMULI** participants.

In summary, we conclude that the SIAM model's state-contingent view may be more useful as a theoretical starting

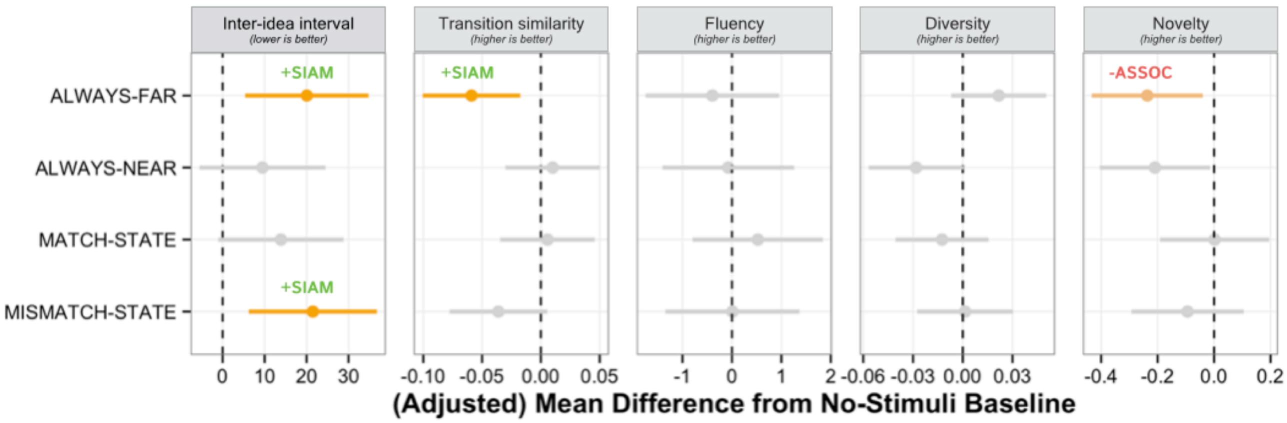

Figure 3. Summary of model-adjusted mean contrasts across dependent measures for each inspiration condition against the NO-STIMULI baseline condition (vertical dashed lines). Mean contrasts are reported on the original scale of the dependent measure. Error bars are 95% confidence intervals. Significant contrasts from the NO-STIMULI baseline (by Dunnett's t-test) are shown in orange. Contrasts that support a theory's prediction are marked green; contrasts that contradict (i.e., go in the opposite direction of) a theory's prediction are marked red. Here, SIAM's predictions for the ALWAYS-FAR (for *inter-idea interval, transition similarity*), ALWAYS-NEAR (for *inter-idea interval, transition similarity, fluency*), and MISMATCH-STATE conditions (for *inter-idea interval*) are supported, while the associationist theory's predictions for ALWAYS-NEAR is contradicted (for *novelty*).

point than the associationist theory of creativity for guiding the design of inspiration delivery systems. In practical terms, our findings suggest that following the associationist recommendation to deliver semantically far inspirations throughout ideation (including during productive ideation, as in the **ALWAYS-FAR** condition) is inadvisable, as it provides uncertain benefits for diversity, and relatively certain costs for novelty and within-category exploration. The theoretical assertions of SIAM suggest that the reduction in novelty arises from disruption of the deep-exploration pathway to novel ideas [11,36,41]. However, it is still unclear if delivering near stimuli during productive ideation and far stimuli during impasses maximizes benefits for ideation. In the limitations section, we suggest methodological changes that might help future studies explore this issue further. We remind the reader that we make no claims about the relative merits of the associationist or SIAM theories for explaining creativity *in general*: we restrict our claims to their relative merits for guiding the design of effective inspiration delivery systems.

Limitations

One limitation of our study is that not all participants entered an impasse state, partially limiting our ability to observe the effects of stimuli during that state. This might be one reason we did not detect an advantage for SIAM's hypothesized best condition (**MATCH-STATE**), since—possibly due to the benefits of receiving near stimuli during productive ideation—many of those participants did not have the opportunity to receive interesting pointers to new areas of exploration (only available in the impasse state in that condition). Perhaps a longer time scale than 8 minutes (to increase the probability that most participants would run out of ideas) or induced breaks in the session would provide a better opportunity to study the effects of stimuli during impasses. Relatedly, a longer time scale might provide a clearer test of the potential cumulative benefits of inspiration during productive ideation (which might come at a slight cognitive cost).

Finally, our sample consisted of MTurk workers paid $6/hr: while many participants expressed enjoyment in the task, low baseline levels of skill/knowledge are still a possibility, which may have suppressed positive effects of inspirations (to the extent that skill/motivation is required to adequately benefit from the inspirations). These are possible reasons that the four treatment conditions did not significantly outperform the **NO-STIMULI** participants on any measure. We therefore urge caution generalizing these results to other settings with a longer time scale and with more skilled/knowledgeable/motivated participants.

Our system also had a relatively slow response time (~1-2s) when retrieving inspirations. While this response time is likely to be acceptable from a usability perspective, the inspirations may not always have arrived before participants started writing down their next idea, potentially adding noise to the intervention. A quicker response time might enable a cleaner test of the potential benefits of near stimuli during productive ideation.

Broader Implications and Future Work

Inspirations should be Delivered at the Right Moments

Our study is consistent with previous work that suggests that potentially helpful inspirations (e.g., analogous ideas from other domains [48], simple hints [28], or diverse ideas of others [45]) are only helpful when delivered under particular circumstances. These findings underscore the importance of considering not just *which* inspirations should be delivered to improve creative ideation, but also *when* they should be delivered. Considering this larger body of findings yields important issues for further research.

One important issue to consider is whether ideators should receive *any* inspiration during productive ideation. Does receiving any stimuli during productive ideation amount to mere disruptive interruptions [2,3,45])? One state-contingent strategy might be to avoid offering any stimuli during productive ideation, and only offer far stimuli during impasses. Our findings are only partially consistent with this view of inspirations: rather than global deficits associated with any stimulation, we observed meaningful theoretically predicted differences between the conditions, finding that far stimuli was harmful, but near stimuli did not lead to statistically significant deficits. The SIAM model of ideation still provides some theoretical reason to doubt this (ideation depends on having "idea elements" to recombine; having more "idea elements" should improve ideation), as does prior studies on the benefits of seeing the ideas of others during collaborative brainstorming in small groups [14]. Future studies that, like our study, vary not just timing but theoretically meaningful variations in the kind of stimuli presented could provide more clarity on this issue.

Another important issue is the implications of a state-contingent view for computational creativity support tools that operate at longer time scales (e.g., weeks, the lifespan of a project) than what we examined in this body of work on timing (i.e., seconds and minutes). At longer time scales, the lines between cognitive states and between real-time inspiration and user-driven search for inspiration might be blurred. For example, how do users get "stuck", or enter a state of creative flow, within the larger context of a project? How might we we appropriately tailor the behavior of the inspiration tools to these states at those time scales (e.g., patent database search engines [30])?

Finally, if accounting for cognitive states of users is important, how might systems effectively detect when users are in particular cognitive states? In this study, we used a partially user-driven approach to detecting users' cognitive states. While this is a reasonable approach for detecting switches to impasse states, future work might explore the use of behavioral markers that predict the onset of an impasse (e.g., slowed inter-idea interval, excessively high inter-idea similarity) and prevent, rather than respond to it. These behavioral markers could be augmented with physiological markers (e.g., the fNIRS signals we obtained in our validation study), to obtain more nuanced and accurate representations of user states. Advances in the portability and wearability of these physiological sensors open up exciting new avenues for designing creativity support systems that respond to users' "implicit input" [47]. This approach could be especially productive to the extent that changes in cognitive states happen more on a continuum than a binary state shift: systems that can detect early/mild stages of impasse and pre-emptively introduce interventions to prevent impasses might represent a new class of creativity support tools that promote extended states of creative "flow" [13].

Far Stimuli Should be Used with Caution

Our findings also have broader implications for the role of semantic distance in creative inspiration. Far (rather than near) stimuli have generally been thought to be more useful for provoking creative mental leaps [19,22,29,40,50]. However, recent investigations are beginning to challenge and refine that claim [10,39,51], pointing out, for example, that overreliance on semantically far stimuli can harm creative performance [10]. Much of this recent work has focused on the impact of inspirations that are far from one's *problem domain* (e.g., drawing inspiration from pendulum motions in grandfather clocks when generating ideas for a new approaches to generating electricity in developing countries). In this work, we extend the notion that far stimuli should be used with caution to the related but distinct notion of semantic distance of inspirations from one's current thinking. We do not mean to argue that far stimuli are unimportant or that they should be avoided entirely; rather, our findings, together with other work on semantic distance, suggest that future research should explore *when* and *how* creators can best take advantage of semantically far inspirations.

Dual Pathways to Creative Outcomes

Finally, our finding that far stimuli during productive ideation not only reduced within-category fluency, but also reduced novelty of ideas (as predicted by SIAM), lends support to the "dual pathway" view of creative ideation [11,36,41], which posits that iteration and deep exploration within categories is an important pathway to creative (not just better quality) ideas, perhaps just as important as creative "mental leaps" to remote regions of a solution space [50], or combining semantically very different ideas [34]. Future research on large-scale collaborative ideation could build on this view to explore how to coordinate the crowd to deeply explore *within* solution approaches, as a complement to promoting cross-pollination of ideas.

CONCLUSION

In this paper, we empirically examined competing theoretical recommendations for how inspirational delivery systems on collaborative ideation platforms should account for semantic distance of inspirational stimuli. In an online ideation experiment, we find that following the associationist theory's recommendation to always provide far stimuli yields uncertain benefits for idea diversity and relatively certain costs for within-category fluency and idea novelty. Our research suggests that far inspirations can be harmful for creativity if delivered during productive ideation, and that collaborative inspiration systems could be improved by accounting for ideators' cognitive states.

ACKNOWLEDGEMENTS

We thank all of the people (both in-person and on MTurk) who participated in the experiment, particularly those who gave valuable feedback on early versions of the experiment. We are also grateful for support from National Science Foundation awards IIS-1122206 and IIS-1122320.

REFERENCES

1. Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. 2015. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (C&C '15), 103–106. https://doi.org/10.1145/2757226.2757252
2. Brian P. Bailey and Shamsi T. Iqbal. 2008. Understanding Changes in Mental Workload During Execution of Goal-directed Tasks and Its Application for Interruption Management. *ACM Trans. Comput.-Hum. Interact.* 14, 4: 21:1–21:28. https://doi.org/10.1145/1314683.1314689
3. Brian P Bailey, Joseph Konstan, John V Carlis, and others. 2000. Measuring the effects of interruptions on task performance in the user interface. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, 757–762.
4. Jonali Baruah and Paul B. Paulus. 2011. Category assignment and relatedness in the group ideation process. *Journal of Experimental Social Psychology* 47, 6: 1070–1077.
5. Roger E. Beaty and Paul J. Silvia. 2012. Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts* 6, 4: 309–319.
6. Osvald M. Bjelland and Robert Chapman Wood. 2008. An Inside View of IBM's' Innovation Jam'. *MIT Sloan management review* 50, 1: 32–40.
7. Joel Chan, Steven C. Dang, and Steven P. Dow. 2016. Improving crowd innovation with expert facilitation. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*. https://doi.org/10.1145/2818048.2820023
8. Joel Chan, Steven Dang, and Steven P. Dow. 2016. IdeaGens: Enabling Expert Facilitation of Crowd Brainstorming. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (CSCW '16 Companion), 13–16. https://doi.org/10.1145/2818052.2874313
9. Joel Chan, Steven Dang, and Steven P. Dow. 2016. Comparing Different Sensemaking Approaches for Large-Scale Ideation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2717–2728. Retrieved May 16, 2016 from http://joelchan.me/files/2016-chi-sensemaking-ideation.pdf
10. Joel Chan, Steven P. Dow, and Christian D. Schunn. 2015. Do The Best Design Ideas (Really) Come From Conceptually Distant Sources Of Inspiration? *Design Studies* 36: 31–58. https://doi.org/10.1016/j.destud.2014.08.001
11. Joel Chan and Christian D. Schunn. 2015. The importance of iteration in creative conceptual combination. *Cognition* 145: 104–115. https://doi.org/10.1016/j.cognition.2015.08.008
12. Allan M. Collins and Elizabeth F. Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82, 6: 407–428.
13. Mihaly Csikszentmihalyi. 1997. *Flow and the Psychology of Discovery and Invention*. HarperPerennial, New York, NY.
14. Darleen M. DeRosa, Carter L. Smith, and Donald A. Hantula. 2007. The medium matters: Mining the long-promised merit of group interaction in creative idea generation tasks in a meta-analysis of the electronic group brainstorming literature. *Computers in Human Behavior* 23, 3: 1549–1581. https://doi.org/10.1016/j.chb.2005.07.003
15. Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2010. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-efficacy. *ACM Trans. Comput.-Hum. Interact.* 17, 4: 18:1–18:24. https://doi.org/10.1145/1879831.1879836
16. Charles W. Dunnett. 1955. A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association* 50, 272: 1096–1121. https://doi.org/10.1080/01621459.1955.10501294
17. Manaal Faruqui and Chris Dyer. 2015. Non-distributional Word Vector Representations. In *Proceedings of ACL*.
18. Katherine Fu, Joel Chan, Christian Schunn, Jonathan Cagan, and Kenneth Kotovsky. 2013. Expert representation of design repository space: A comparison to and validation of algorithmic output. *Design Studies* 34, 6: 729–762. https://doi.org/10.1016/j.destud.2013.06.002
19. Dedre Gentner and Arthur B. Markman. 1997. Structure mapping in analogy and similarity. *American Psychologist* 52, 1: 45–56.
20. Joy P. Guilford. 1950. Creativity. *American Psychologist* 5: 444–454.
21. Nitin Gupta, Yoonhee Jang, Sara C. Mednick, and David E. Huber. 2012. The Road Not Taken: Creative Solutions Require Avoidance of High-Frequency Responses. *Psychological Science*.
22. Keith J. Holyoak and Paul Thagard. 1996. *Mental leaps: Analogy in creative thought*. Cambridge, MA.
23. Seth Hunter and Pattie Maes. 2008. WordPlay: A table-top interface for collaborative brainstorming and decision making. *Proceedings of IEEE Tabletops and Interactive Surfaces*: 2–5.
24. Theodore J. Huppert, Solomon G. Diamond, Maria A. Franceschini, and David A. Boas. 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics* 48, 10: D280. https://doi.org/10.1364/AO.48.00D280
25. Theodore J. Huppert, Rick D. Hoge, Solomon G. Diamond, Maria A. Franceschini, and David A. Boas.

2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage* 29, 2: 368–382. https://doi.org/10.1016/j.neuroimage.2005.08.065

26. Elahe Javadi and Wai-Tat -. T. Fu. 2011. Idea Visibility, Information Diversity, and Idea Integration in Electronic Brainstorming. In *Proceedings of the 6th International Conference on Foundations of Augmented Cognition: Directing the Future of Adaptive Systems* (FAC'11), 517–524.
27. Elahe Javadi, Joseph Mahoney, and Judith Gebauer. 2013. The impact of user interface design on idea integration in electronic brainstorming: an attention-based view. *Journal of the Association for Information Systems* 14, 1: 1–21.
28. Craig A. Kaplan and Herbert A. Simon. 1990. In search of insight. *Cognitive Psychology* 22, 3: 374–419.
29. Stefan W. Knoll and Graham Horton. 2011. The Impact of Stimuli Characteristics on the Ideation Process: An Evaluation of the Change of Perspective "Analogy." In *Proceedings of the 44th Hawaii International Conference on System Sciences*.
30. S. Koch, H. Bosch, M. Giereth, and T. Ertl. 2009. Iterative Integration of Visual Insights during Scalable Patent Search and Analysis. In *Proceedings of IEEE Transactions on Visualization and Computer Graphics*.
31. Arthur Koestler. 1964. *The act of creation*. Macmillan, Oxford, England.
32. T. K. Landauer and S. T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104, 2: 211–240.
33. Brian Lee, Savil Srivastava, Ranjitha Kumar, Ronen Brafman, and Scott R. Klemmer. 2010. Designing with Interactive Example Galleries. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 2257–2266. https://doi.org/10.1145/1753326.1753667
34. Sara A. Mednick. 1962. The associative basis of the creative process. *Psychological Review* 69, 3: 220–232.
35. Ana Sofia Morais, Henrik Olsson, and Lael J. Schooler. 2012. Mapping the Structure of Semantic Memory. *Cognitive Science*. https://doi.org/10.1111/cogs.12013
36. Bernard A. Nijstad, Carsten K. W. De Dreu, Eric F. Rietzschel, and Matthijs Baas. 2010. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology* 21: 34–77. https://doi.org/10.1080/10463281003765323
37. 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. https://doi.org/10.1207/s15327957pspr1003_1
38. Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)* 12: 1532–1543.
39. David N. Perkins. 1983. Novel Remote Analogies Seldom Contribute to Discovery. *The Journal of Creative Behavior* 17, 4: 223–239.
40. Tony Poze. 1983. Analogical connections: The essence of creativity. *The Journal of creative behavior* 17, 4: 240–258.
41. Eric F. Rietzschel, Bernard A. Nijstad, and Wolfgang Stroebe. 2007. 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: 933–946. https://doi.org/10.1016/j.jesp.2006.10.014
42. Albert Rothenberg. 1979. *The emerging goddess: The creative process in art, science, and other fields*. University of Chicago Press Chicago.
43. 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. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing*. https://doi.org/10.1145/2675133.2675239
44. 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. In *Proceedings of 2016 ACM Conference on User Interface Software and Technology (UIST 2016)*.
45. Pao Siangliulue, Joel Chan, Krzysztof Gajos, and Steven P. Dow. 2015. Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the ACM Conference on Creativity and Cognition*. https://doi.org/10.1145/2757226.2757230
46. Erin Treacy Solovey, Francine Lalooses, Krysta Chauncey, Douglas Weaver, Margarita Parasi, Matthias Scheutz, Angelo Sassaroli, Sergio Fantini, Paul Schermerhorn, Audrey Girouard, and Robert J.K. Jacob. 2011. Sensing Cognitive Multitasking for a Brain-based Adaptive User Interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), 383–392. https://doi.org/10.1145/1978942.1978997
47. Erin Treacy Solovey, Daniel Afergan, Evan M. Peck, Samuel W. Hincks, and Robert J. K. Jacob. 2015. Designing Implicit Interfaces for Physiological Computing: Guidelines and Lessons Learned Using fNIRS. *ACM Trans. Comput.-Hum. Interact.* 21, 6: 35:1–35:27. https://doi.org/10.1145/2687926
48. Ian Tseng, Jarrod Moss, Jonathan Cagan, and Kenneth Kotovsky. 2008. The role of timing and analogical similarity in the stimulation of idea generation in design. *Design Studies* 29, 3: 203–221.

49. Arno Villringer and Britton Chance. 1997. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences* 20, 10: 435–442. https://doi.org/10.1016/S0166-2236(97)01132-6
50. Thomas B. Ward. 1998. Analogical distance and purpose in creative thought: Mental leaps versus mental hops. In *Advances in Analogy Research: Integration of Theory and Data from the Cognitive, Computational, and Neural Sciences*, Keith J. Holyoak, Dedre Gentner and B. Kokinov (eds.). Sofia, Bulgaria, 221–230.
51. R. W. Weisberg. 2009. On "out-of-the-box" thinking in creativity. In *Tools for innovation*, A. B. Markman and K. L. Wood (eds.). New York, NY, 23–47.