Early and Repeated Exposure to Examples Improves Creative Work

Chinmay Kulkarni
Stanford University HCI Group
Stanford, CA 94305-9035 USA
chinmay@cs.stanford.edu

Steven P. Dow
Carnegie Mellon, HCI Institute
Pittsburgh, PA 15213
spdow@cs.cmu.edu

Scott R Klemmer
Stanford University HCI Group
Stanford, CA 94305-9035 USA
srk@cs.stanford.edu

Abstract

This article presents the results of an online creativity experiment ($N = 81$) that examines the effect of example timing on creative output. In the between-subjects experiment, participants drew animals to inhabit an alien Earth-like planet while being exposed to examples early, late, or repeatedly during the experiment. We find that exposure to examples increases conformity. Early exposure to examples improves creativity (measured by the number of common and novel features in drawings, and subjective ratings by independent raters). Repeated exposure to examples interspersed with prototyping leads to even better results. However, late exposure to examples increases conformity, but does not improve creativity.

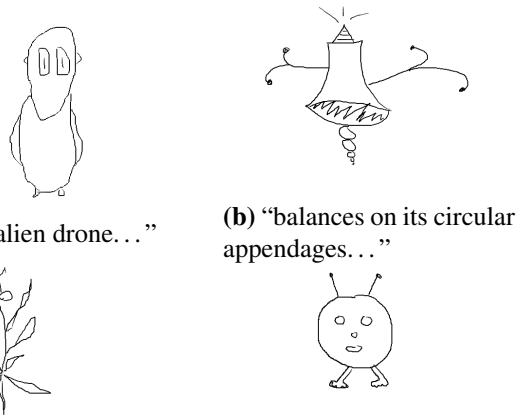
Introduction

Examples are considered "a cornerstone of creative practice" (Herring et al., 2009). Leveraging examples of prior work is an established technique in design (Buxton & Buxton, 2007), and many design programs encourage students to use examples of existing designs (Schön, 1985). However, the strategies employed by designers to seek and use examples is largely ad-hoc (Newman & Landay, 2000).

Frequently, these strategies differ in timing—in an informal survey we conducted among designers around Stanford University, one respondent described inspirational examples as "huge parts of my initial steps. I need to know as much as I can about the topic before I feel comfortable moving forward." In contrast, another said that "I don't do this [look at examples] at the very beginning because it gets your mind stuck in one way of thinking." This fear of conformity was echoed by other participants, and one went on to say that he looked for inspiration only when "facing a creative block."

These different strategies suggest that examples may modify the creative process differently depending on the point in the design process at which they are presented. This leads to the practical question: what are the tradeoffs of looking at examples earlier or later in the design process? Furthermore, even if there is an "ideal" time to view examples, some designers feel ubiquitous information access and their own "thirst for knowledge" bombards them constantly with examples (Herring et al., 2009). How does this repeated exposure to examples affect the creative process?

This article presents the results of an online creativity experiment we conducted on Amazon Mechanical Turk. Participants in the experiment generated drawings of alien creatures as a creative task. The pervasive use of sketches to develop and communicate conceptual designs in the creative fields (Suwa & Tversky, 1997), and the use of similar tasks in prior work Ward (1994) inspired the choice of the drawing task. Focusing on drawings of alien figures makes this task readily accessible to non-designers (see Figure 1 for a sampling of drawings created by participants). Participants were randomly assigned to one of four conditions: examples early, examples late, examples early and late, or a control condition without examples. This study's creativity measures were the number of uncommon and novel features in the drawings and Likert-scale ratings by condition-blind raters. Conformity was measured by the number of critical features (features that were directly copied from examples).



(a) "alien drone…"

(b) "balances on its circular appendages…"

(c) "buglike alien…"

(d): "This round furry creature gets around either by walking on his retractable legs or rolling across surfaces…"

Figure 1: Sampling of drawings created in our experiment, with excerpts of participant-provided descriptions

This paper's experimental results suggest that while exposure to examples increases conformity, such exposure early in the creative process improves the creativity in the output, while later exposure provides no such benefit. Furthermore, exposure to examples followed by prototyping and subsequent re-exposure to the same examples improved creative output even more. This finding may allay some fears of example bombardment. Lastly, in our experiment, participants exposed to examples created fewer drawings, so these example driven quality improvements may come at the cost of a lower quantity of creative work.

Related work

Examples: Bringing existing solutions to mind is crucial for creative generation (Smith et al., 1993). The Structured Imag-

ination theory by Ward (1994) describes creativity as a multi-step process: in the *recall* step, people bring to mind existing solutions and constructs. Then, in the *modification* step, these constructs are altered in novel ways. Similar analogical processes are found in other areas of cognition such as analysis and learning (Gentner & Colhoun, 2010).

Designers often incorporate features from examples directly into their work (Marsh & Bower, 1993, "inadvertent plagiarism"); but examples also "ultimately alter the nature of the creative product" in more subtle ways (Marsh et al., 1996). Lee et al. (2010) found that designing with examples generally improves the quality of creative work. These findings have also led to tools for discovering, storing and retrieving examples (Kerne et al., 2008; Ritchie et al., 2011).

The current work is an extension of Marsh et al. (1996) (itself an extension of Smith et al., 1993), so we describe Marsh et al.'s experiment in more detail. In their experiment, participants generated drawings of non-Earth-like creatures to inhabit an alien planet similar to Earth. In the example conditions, experimenters provided participants example drawings of aliens at the start of the experiment. Example drawings all had certain attributes, or *critical* features, in common—four legs, antennae and a tail. The proportion of these critical features incorporated into participants' own drawings was used as a measure of conformity. The proportion of other, non-critical, features was used as a measure of creativity. These non-critical features were classified as either *novel* (not commonly found on animals, such as speakers or propellors), *uncommon* (such as a pouch or tentacles), or *common* (such as a nose, mouth or two legs).

Participants exposed to examples incorporated more critical features in their drawings, but not at the expense of novel and uncommon features. Instead, their drawings contained fewer *common* features. This suggests that while examples increase conformity by increasing activation of critical features, they do not block retrieval of original ideas (such as novel and uncommon features).

We use Marsh et al.'s feature-based evaluation metric, and extend their work by examining how the example timing affects creative output. In addition, we study the effects of repeated exposure to examples in the creative process.

Research methods Our experiment uses a task (drawing sketches of alien figures) that has previously been employed to study creativity in a context of no prior training (Marsh & Bower, 1993; Ward, 1994). Drawing tasks have also been demonstrated to be appropriate for online experiments (Yu & Nickerson, 2011).

This experiment was run on Amazon Mechanical Turk (`www.mturk.com`), a web-based crowdsourcing plaform. This platform has been used for experiments on affect and creativity (Lewis et al., 2011). Mechanical Turk workers have also been employed to provide perception responses (Heer & Bostock, 2010), objective labels (Deng et al., 2009; Snow et al., 2008), and subjective ratings (Dow et al., 2011).

Condition	Task before first session	Task before second session
Control	Think	Think
Early	Examples	Think
Late	Think	Examples
Repeated	Examples	Examples

Table 1: Experimental conditions

Experiment

Our experiment had two goals. First, we wanted to see if exposure to examples at the start of a creative process leads to a different quality of creative output in contrast to exposure when the creative process is underway. Second, we wanted to to investigate the role of repeated exposure to examples.

Our initial hypothesis was that exposure to examples later in the creative process would have the same creative benefits but lower conformity than exposure at the start. This hypothesis was motivated by Weisberg (1999), who observed that creative failures are more often explained by the absence of relevant information than the presence of irrelevant information. Furthermore, the presence of one's own ideas would inhibit the adoption of sub-optimal ideas from late exposure to examples (mirroring the intuitions of some designers).

In the case of repeated exposure, the activation account would predict that, similar to showing more examples at once, this would result in greater degree of conformity due to higher activation of features present in examples.

Participants

We solicited US-resident participants on Mechanical Turk with a compensation of US$1.00. 81 participants responded (27 male, 54 female; median age 34). All participants reported a high-school diploma or a higher degree. This between-subjects experiment randomly assigned participants to one of four conditions.

Procedure

The experiment comprised two drawing sessions, each lasting 7 minutes. Participants were asked to create as many drawings as they could during the drawing session. To encourage this (and discourage participants from spending time perfecting only a few drawings), the experimental platform included a clear-canvas tool but no line-eraser tool (Figure 2).

Each session was preceded by a condition-specific task in which participants were either exposed to examples, or asked to think about the aliens they planned to draw in the next session (Table 1).

At the start of the experiment, all participants saw a Web page with instructions adapted from Marsh et al. (1996) to account for two drawing sessions and a break (see `hci.stanford.edu/example/aliens` for actual prompts used).

For the **Example** task, participants were shown three example alien drawings for 90 seconds (see Figure 4). We used drawings from (Marsh et al., 1996), p. 672 . Using the prompt

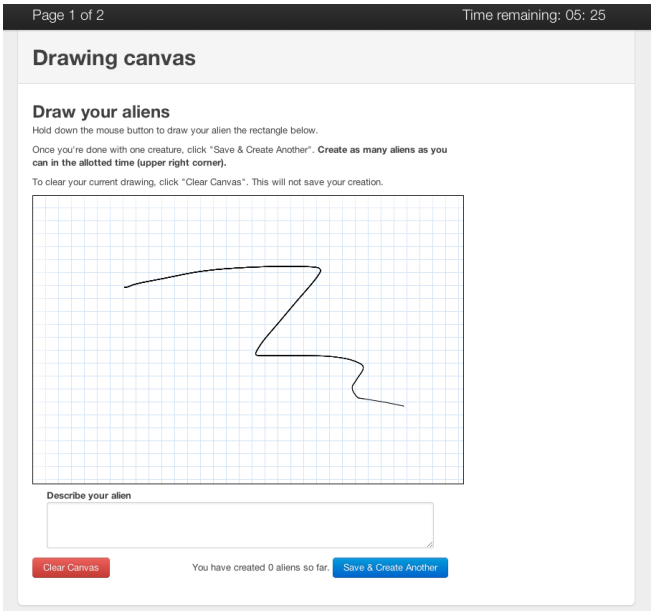


Figure 2: Drawing canvas with time remaining (top right), and an option to clear the canvas (bottom left, red)

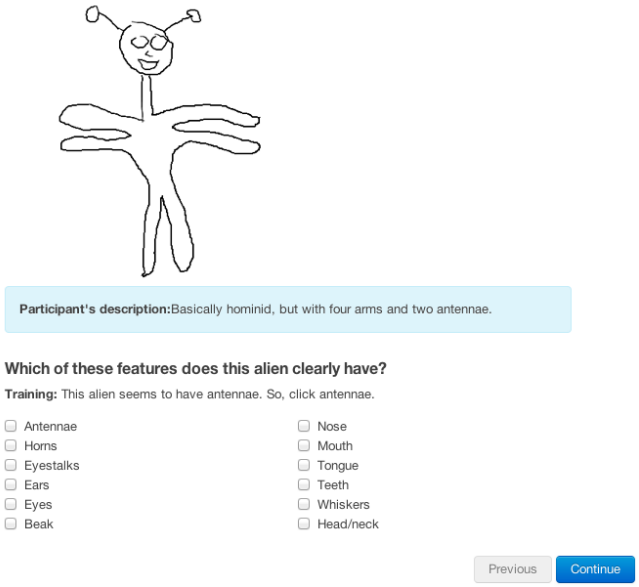


Figure 3: Training interface for annotators. The training interface shows what features to label ("click antennae"). The actual annotation is performed on an identical list of features.

of Marsh et al. (1996) (and Smith et al., 1993), participants were instructed that examples were only shown to help them create their original creations, and that we did not want them to copy the examples in any aspect. For the **Think** task, participants were asked to "think about aliens" they planned to draw in the next session for 90 seconds. In the Repeated Examples condition, participants saw the same three examples before both drawing sessions.

After the second drawing session, participants filled out a survey that covered demographics, artistic interest and ability and the thought-process they followed while drawing.

Labeling features in drawings

Participants generated a total of 543 drawings. Each drawing was labelled with the features it incorporated from the feature set of Marsh et al. (1996) (Appendix). Drawings were annotated on Mechanical Turk, since the features were well-defined. All workers were US resident and at least 18 years of age, and were compensated US$0.50 for the task. Workers who participated in the experiment were disallowed from the annotation task (and vice-versa). All annotators were blind to experimental condition.

Workers were trained using a drawing from a pilot participant (Figure 3). Then, each worker annotated a set of seven randomly assigned drawings. Workers also rated how creative they found the drawing on a 7-point Likert scale (each annotator saw at least one drawing from each condition). Lastly, annotators could flag offensive (or non-alien) drawings. Upon review, 34 flagged drawings were discarded by the authors. Each drawing was annotated by two workers. Disputes in annotation were resolved by the authors.

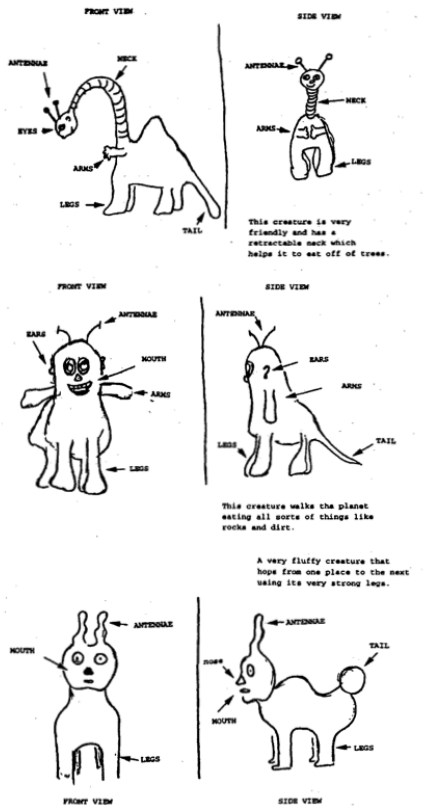


Figure 4: Example drawings provided to participants. All examples contain critical features—four legs, antennae, and a tail.

Condition	Critical	Common	Uncommon	Novel	Total	Drawings per session	Likert Rating
Control	0.39	4.21	0.95	0.47	6.03	4.00	3.71
Early	**0.57**	3.91	*1.15*	0.40	6.04	**3.00**	**4.10**
Late	**0.52**	*3.82*	0.78	0.45	*5.57*	**3.68**	3.43
Repeated	**0.64**	4.20	**1.21**	0.54	6.60	**3.00**	**4.22**

Table 2: Table of means. Means that differed from control at $p < 0.05$ are **bold**, those marginally significant ($p < 0.1$) are in *italics* (p-values from the post-hoc analysis using mixed models, see section Results).

Results

We analyze data using a mixed-effects linear model. Since participants drew multiple drawings per drawing session, unless noted, we consider the participant as a random effect with a fixed intercept; and experimental condition, drawing session (first or second), and an interaction term as fixed-effects in all our analysis below. Reported p-values are from a Monte-Carlo (MCMC) simulation (Baayen et al., 2008).

Examples increase conformity

Following Smith et al. (1993), conformity was measured as the number of critical features incorporated per drawing. Without controlling for the drawing session, examples shown at the start of the experiment increased the number of critical features that were incorporated into drawings ($t(507) = 2.06$, $p < 0.05$), consistent with results from (Smith et al., 1993; Marsh et al., 1996). Participants in the Late Examples condition show higher conformity in the second drawing session (*i.e.* post-exposure) [$t(419) = 1.83$, $p = 0.07$].

Early exposure increases uncommon features

The number of uncommon features per drawing increased in the Early Examples condition ($t(419) = 1.61, p = 0.06$), and in the Repeated Examples condition ($t(419) = 1.72, p < 0.05$), but not in the Late Examples condition ($t(419) = -0.45, p = 0.649$) (Figure 5). The number of novel features did not vary significantly across condition. Participants in the Late exposure condition created drawings with marginally fewer common features($t(419) = -1.33, p = 0.09$) and fewer total number of features ($t(419) = -1.30, p = 0.09$).

Early and Repeated exposure leads to higher subjective ratings

Annotators rated drawings in the Early Examples and the Repeated Examples conditions higher ($t = 2.24, p < 0.05$ and $t = 2.65, p < 0.01$, respectively). Intra-class correlation amongst raters (average, random raters) was 0.54 ($F(508, 508) = 2.2, p < 0.001$).

Examples reduce number of drawings

Unlike Marsh et al. (1996), participants created fewer drawings per session in all example conditions[1] [Early: $t(149) = -2.50, p < 0.05$; Late: $t(149) = -2.14$, $p < 0.05$, Repeated: $t(149) = -2.63$, $p < 0.05$] (Figure 6). Participants in the Late Examples condition created fewer drawings after exposure to examples ($\mu_{before} = 4.10, \mu_{after} = 3.13, t(149) = 1.91$, $p_{interaction} < 0.05$).

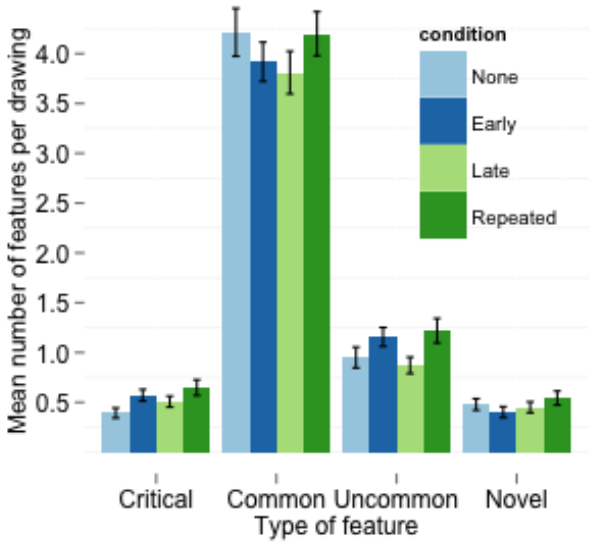


Figure 5: Participants in early and repeated exposure conditions included more uncommon features compared to Late exposure/control conditions.

Discussion

Example timing affects creative output

These results suggest that exposure to examples at any time increases conformity. However, early exposure increases the number of uncommon features and subjective ratings of creativity, while late exposure provides no such benefits. This runs counter to both our initial hypothesis and the intuitions of many designers who delay looking at examples in an effort to reduce fixation and think "out of the box" (Jansson & Smith, 1991).

One possible explanation for these effects is that early exposure to examples aids the designer in understanding the scope of acceptable solutions to a problem, and helps form an initial representation of the creative concept (Heit, 1992). Prototyping results in subsequent abstraction and refinement of the initial representation (Lim et al., 2008). Without initial exposure to examples, the refined representation may dif-

[1]Since the number of drawings is not a repeated measure, analysis uses a fixed-effects model with interaction, the experimental condition and the type of session being independent variables.

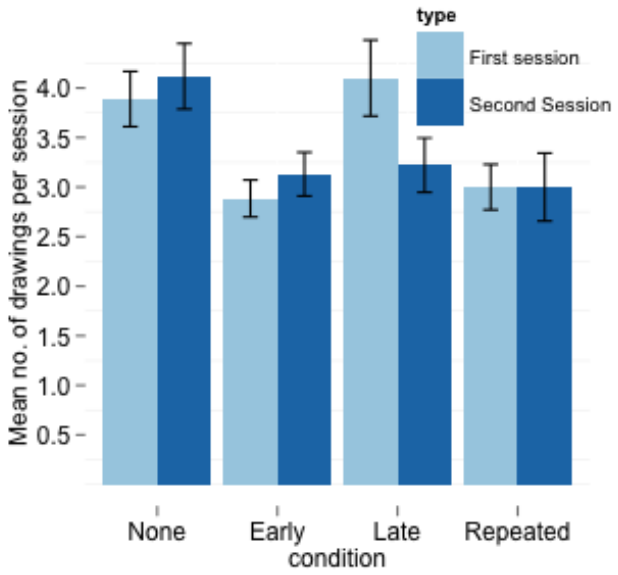


Figure 6: Participants drew fewer drawings when examples were shown (in the Late examples condition, participants drew fewer drawings in the second session).

fer widely from the one embodied in examples, which would make it harder to map concepts from the example to one's own representation. When exposure is only for a short duration (90s in our experiment), it is possible that only concepts with high enough activation, such as critical features in our experiment, are transferred (motivated by Boroditsky, 2007).

Another counter-intuitive experimental result is that repeated exposure to the same examples led to higher creative quality. This may also be explained by a seeding-and-transfer account. Initial exposure to examples prevents the refined representation formed by prototyping from diverging greatly from the one embodied in the examples. This refined yet similar representation would then allow the designer to learn different concepts on re-exposure to the same example.

In essence, the crucial ingredient that allows repeated exposure to improve creativity might be the prototyping that occurs between exposures.

Why did examples yield fewer drawings?

Examples play a dual role in design– first, they inspire different solutions and ways of thinking. Second, they help form expectations about what characteristics a solution needs to have (Herring et al., 2009). The decrease in in the number of drawings created may be due to this second role. Seeing examples may have signaled a higher threshold for "acceptable" drawings, resulting in participants spending more time on each drawing, and creating fewer drawings overall.

Our data suggest that this expectation-setting role has a different behavior than the inspirational role. While the number of drawings created decreased nearly uniformly post-exposure, changes in creativity measures (uncommon features and subjective ratings) were non-uniform. Therefore, while examples may set expectations any time they are presented (including late in the design process), their inspirational value may be time-dependent.

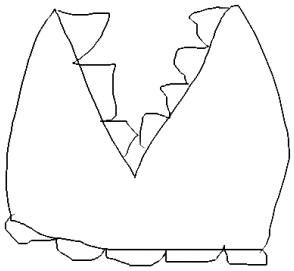


Figure 7: Participant-provided description: "An ambush predator that does move very much, but lures prey into its mouth using scent to make them think there is food there. It only occasionally shifts using the pads on its bottom, which can also suck up nutrients from the ground or water for emergencies." Rated highly creative by our raters, this drawing has no novel or uncommon visual features, and uses a non-visual feature (scent).

Multiple Measures of Creativity

Results from both the feature-counting measure of creativity from prior work and the Likert-scale ratings provided by annotators are largely consistent. While the Likert ratings are subjective, they better capture the creativity in some drawings that combine common or critical features in a novel way, or use a non-visual feature (for *e.g.* see Figure 7). Using both together provides a better characterization of creativity.

Conclusions and future work

This work demonstrates the benefits of early and repeated exposure to examples on creative work. In addition, it suggests that conformity may be the price one pays for these gains, regardless of when examples are seen. Hopefully, these results will encourage designers to seek examples early and often in the design process, when they are most useful.

This experiment also demonstrates a replication (and extension) of creativity studies in an online crowdsourced environment. Crowdsourced experiments often offer a lower cost, have a faster time to completion, and provide access to wider populations (Heer & Bostock, 2010). This paper's experiment and the labeling tasks took one week on Amazon Mechanical Turk. This was possible because the labeling scheme from Marsh et al. (1996) provided this study with a clear taxonomy of features that could be easily labeled by non-experts. We suggest crowdsourcing as a viable platform both for experiments that do not need participants with specialized skills or background (or modification per participant) and for analysis/labeling tasks that easily verifiable.

This work also raises a number of questions. First, the results of the repeated-exposure experimental condition indicate that the processes of prototyping and learning from examples may be intertwined in a creative task. Further empir-

ical studies could characterize the precise nature of this interaction. Second, this work shows that repeated exposure to examples is beneficial. How does the frequency of (or interval between) such exposures affect this result? Third, designers often spend years acquiring skills and specific domain knowledge. How do such skills and knowledge affect their interaction with examples? Furthermore, similar to cross-cultural effects of prototyping (Kim & Hinds, 2012), are effects of examples different in different cultures? Finally, how can the results of this work inform the design of tools that support creative work?

Acknowledgements

We thank the Hasso Plattner Design Thinking Research Program for supporting this work.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390–412.

Boroditsky, L. (2007). Comparison and the development of knowledge. *Cognition*, *102*(1), 118–128.

Buxton, B., & Buxton, W. (2007). *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conf on* (pp. 248–255).

Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D., & Klemmer, S. (2011). Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proc of CHI: ACM conf. on human factors in computing systems* (pp. 2807–2816).

Gentner, D., & Colhoun, J. (2010). Analogical processes in human thinking and learning. *Towards a Theory of Thinking*, 35–48.

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proc of CHI: ACM conf. on human factors in computing systems* (pp. 203–212).

Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, *24*(3), 341–380.

Herring, S., Chang, C., Krantzler, J., & Bailey, B. (2009). Getting inspired!: understanding how and why examples are used in creative design practice. In *Proc of CHI: ACM conf. on human factors in computing systems* (pp. 87–96).

Jansson, D., & Smith, S. (1991). Design fixation. *Design Studies*, *12*(1), 3–11.

Kerne, A., Koh, E., Smith, S., Webb, A., & Dworaczyk, B. (2008). combinformation: Mixed-initiative composition of image and text surrogates promotes information discovery. *ACM Trans on Information Systems (TOIS)*, *27*(1), 5.

Kim, H., & Hinds, P. (2012). Harmony vs. disruption: The effect of iterative prototyping on teams creative processes and outcomes in the west and the east. In *Proc. icic: International conf. on intercultural collaboration.* ACM.

Lee, B., Srivastava, S., Kumar, R., Brafman, R., & Klemmer, S. (2010). Designing with interactive example galleries. In *Proc of CHI: ACM conf. on human factors in computing systems* (pp. 2257–2266).

Lewis, S., Dontcheva, M., & Gerber, E. (2011). Affective computational priming and creativity. In *Proc of CHI: ACM conf. on human factors in computing systems* (pp. 735–744).

Lim, Y., Stolterman, E., & Tenenberg, J. (2008). The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *15*(2), 7.

Marsh, R., & Bower, G. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 673.

Marsh, R., Landau, J., & Hicks, J. (1996). How examples may (and may not) constrain creativity. *Memory & Cognition*.

Newman, M., & Landay, J. (2000). Sitemaps, storyboards, and specifications: a sketch of web site design practice. In *Proc of DIS: ACM conf. on designing interactive systems* (pp. 263–274).

Ritchie, D., Kejriwal, A., & Klemmer, S. (2011). d. tour: style-based exploration of design example galleries. In *Proc. of UIST: ACM symposium on user interface software and technology* (pp. 165–174).

Schön, D. (1985). *The design studio: An exploration of its traditions and potentials*. RIBA Publications for RIBA Building Industry Trust London.

Smith, S., Ward, T., & Schumacher, J. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, *21*(6), 837–845.

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. (2008). Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proc of the conf. on empirical methods in natural language processing* (pp. 254–263).

Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? a protocol analysis. *Design studies*, *18*(4), 385–403.

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

Weisberg, R. (1999). Creativity and knowledge: A challenge to theories. *Handbook of creativity*, 226.

Yu, L., & Nickerson, J. (2011). Cooks or cobblers?: crowd creativity through combination. In *Proc of CHI: ACM conf. on human factors in computing systems* (pp. 1393–1402).