Schema and Metadata Guide the Collective Generation of Relevant and Diverse Work

Xiaotong (Tone) Xu,[1] Judith E. Fan,[2] Steven P. Dow[3]
[1,3]Department of Cognitive Science, [2]Department of Psychology
University of California - San Diego, La Jolla, California USA
{xt, jefan, spdow}@ucsd.edu

Abstract

While most crowd work seeks consistent answers, creative domains often seek more diverse input. The typical crowd mechanisms for controlling quality may stifle creativity, yet removing them altogether could just produce noise. Schemas and metadata provide two mechanisms for embedding existing knowledge into task environments. Schemas are expert-derived patterns designed to structure how people think through a problem. Metadata, on the other hand, illustrate a range of creative input that fits within the structure of a schema. To understand the relative effects of schemas and metadata, we conducted a study where crowd workers are asked to generate creative interpretations for a set of placemaking examples. Crowd workers were guided either by schema plus metadata, schema alone, or neither. We found that showing schema along with crowd-produced metadata helped workers contribute interpretations that are both more on-topic and diverse, compared to using the schema alone or no schema. We discuss the implications on how crowds can creatively build on insights shared by others.

Introduction

Many researchers have explored how to get crowds to do complex and creative work (Chou and Tversky 2020; Chilton et al. 2014). One strategy has been to create workflows where crowd workers ”build on” batches of work done by prior workers (Chilton et al. 2014; Kittur et al. 2011). Workers can get inspiration from prior examples of creative work (Kulkarni, Dow, and Klemmer 2014; Yu and Nickerson 2011). Workers can also enhance prior creative work by assessing, comparing, or combining ideas (Goucher-Lambert and Cagan 2019; Girotto, Walker, and Burleson 2017; Lee et al. 2016; Chiang, Kasunic, and Savage 2018), effectively building on the examples by producing additional data around the creative work. Similarly, crowd workers can build on prior sensemaking by adding, filtering, and interpreting data such that produces an overall synthesis (Zhang, Verou, and Karger 2017; Chilton et al. 2013; 2014; André, Kittur, and Dow 2014). What general strategies can guide the collective generation of diverse and relevant insights for a creative domain? On many web-based platforms, this process of building on prior user contributions happens organically through commenting and tagging (Chilton et al. 2013; Girotto, Walker, and Burleson 2017). For example, on Pinterest, some members share examples of creative work, while other members add comments. These peer-generated interpretations could serve to highlight the salient features, values, or dimensions within the examples. This research explores two key strategies for guiding subsequent interpretations: schemas and metadata.

In the context of crowd work, *schemas* are the mental structures that people form in order to effectively perform tasks (Brewer 1987). Critically, they provide a cognitive framework to help people focus on a reduced set of dimensions, rather than be overwhelmed by the large number of possibilities (Yilmaz et al. 2016). However, the schemas that people use are often implicit and thus their effects on downstream creative work are both challenging to study, and potentially diminished by not being explicit. Towards addressing this issue, prior work has developed techniques for inferring what schemas people use, and explored how making these schemas more explicit improves information foraging work by subsequent workers (Kittur et al. 2014). However, prior work has shown mixed effects how schemas influence creative work. On one hand, schemas might hinder creative work by yielding more consistent behavior across different people (Chilton et al. 2013). On the other hand, schemas may facilitate creative work by providing people with a structured approach for considering diverse possibilities (Liu and Bagrow 2017).

A key complementary strategy available to crowd platforms is to gather *metadata* or information about other data (Barber 2018). For instance, comments posted by Pinterest community members contribute metadata by offering unique insights and perspectives. While generating metadata can be helpful for searching and browsing a collection of information, they can also be difficult to leverage when left unstructured or hidden from view. Schema and metadata work in conjunction, since schema can provide an organizational structure for metadata, and they can accumulate over time, as each new worker contributes their own schema attributes and/or metadata.

To understand the relative value of schema and metadata in crowd platforms, we conducted a study where participants provided creative interpretations about design examples. These design examples varied with respect to several key attributes (i.e., stakeholders, artifacts, context, process, goals). These interpretations were either generated in a completely unconstrained manner or with respect to each of the attributes in this schema. Further, some participants who were provided with the schema built on existing metadata (i.e., prior participants' interpretations). We used text-based natural language processing to measure the relevance (i.e., semantic similarity) and the novelty (i.e., reduced word overlap) of the resulting interpretations in each setting. Our results suggest that providing participants with an explicit schema helped them to produce more relevant interpretations, relative to providing no explicit schema. Moreover, participants who were provided with both a schema and metadata produced more novel interpretations, as measured by decreased reuse of language from the example description, relative to those who were provided with a schema but no metadata. Taken together, these findings suggest that using a combination of both schemas produced by experts and metadata produced by other crowd workers helps them to make contributions that are both useful and diverse. Such findings may have implications for designing platforms that better support the collective generation of insights.

Method

To investigate the effects of schema and metadata, we created an experiment that simulates an asynchronous collaborative design task. Participants reviewed a series of design examples with different stimuli (i.e., the availability of schema and metadata) and offered their own interpretations on the examples, before finally proposing their own proposed solution. We measured the diversity and relevance of participants' interpretations, as well as the quality of their proposed solutions.

Participants

92 workers from Amazon Mechanical Turk were paid $4 to participate in a 30-minute study. All workers were fluent in English and had U.S.-based IP addresses.

Stimuli

Task Domain and Examples Participants were asked to generate a *placemaking* proposal. Placemaking is an urban design strategy to "collectively reimagine and reinvent public spaces" (Schneekloth and Shibley 1995). We chose placemaking as the task domain because: 1) it is complex and multi-faceted, 2) yet accessible to non-experts, and 3) there are many resources for placemaking examples on the Web, including peer-generated interpretations on these examples. We selected ten diverse placemaking examples from a non-profit Pinterest account (Project for Public Spaces 2012) that showcases various ways to transform public spaces to attract residents and tourists. Each example included an image and a short text description.

Schema and Metadata We used a data-driven approach to generate a shared schema and metadata for each placemaking example. First, we recruited N=43 pilot participants to provide open-ended interpretations about

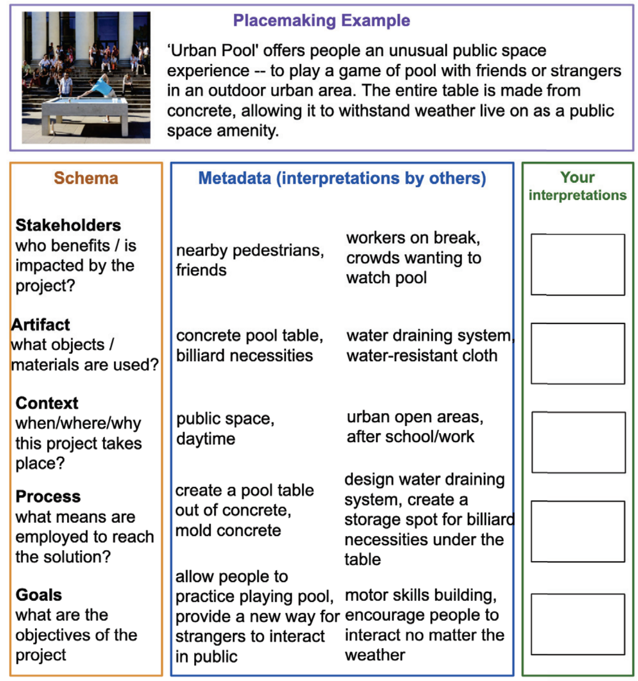

Figure 1: Study task: Participants were asked to write five of their own interpretations of each placemaking example. In the *schema+metadata* condition, participants saw five schema questions, as well as previous interpretations by others (metadata). *Schema* participants only saw the schema questions and did not see the metadata, and *baseline* participants did not see the schema or the metadata.

important design dimensions represented in the placemaking examples. Then, our research team conducted an affinity-diagramming session to derive five key design dimensions for the schema: 1) *Stakeholders* who benefits/is impacted by the project? 2) *Artifacts* what objects/materials are used? 3) *Context* when/where/why this project takes place? 4) *Process* what means are employed to reach the solution? and 5) *Goals* what are the objectives of the project? These five dimensions were intended to help people make sense of placemaking examples, but they also potentially align with other design domains (Studer et al. 2018).

Second, to collect metadata, we administered a questionnaire to N=7 pilot participants who viewed all ten placemaking examples, and provided interpretations with respect to each dimension in the schema. We then selected four interpretations for each example along each dimension giving us a total of 200 interpretations. The four interpretations on each dimension of each example were selected to ensure breadth, and were lightly edited for grammar.

Procedure

Participants filled out a consent form and then read an overview of the study and design brief. The design brief gave background information about a Midwestern US town in a recession and in need of innovation to revitalize public

Baseline	Schema	Schema+Metadata
durable	tourists	promotes friendship among strangers
interactive	table concrete	pool cues and balls
fun	museum, park	can be used during lunch time as stress relief
activity	use any concrete	perhaps light the area for night time use
socialization	socializing	provides a sense of friendship among strangers

Table 1: Sample interpretations on the 'Urban Pool' placemaking example written by participants in each condition.

space and to attract tourists. Participants were told they would review a series of placemaking examples, produced interpretations on these examples, and then invent their own solution. Their proposed solution would need to cost less than $10k, could not require heavy infrastructure, and could not exceed a one-year construction timeline. Participants were informed their solution would be judged based on its novelty and usefulness.

Each participant was randomly assigned to one of three conditions: In the **schema+metadata** condition, participants can see the five questions associated with the schema, as well as previous interpretations by other people (see Figure 1), while they viewed each of the ten placemaking examples. In the **schema** condition, participants see the five schema questions, but no interpretations from others. In the **baseline** condition, participants see no questions or prior interpretations; they are only asked for their own insights on the examples.

The ten placemaking examples appeared in random order and included an image and text description. After reviewing the example, participants clicked a 'next' button, which revealed condition-specific stimuli (schema+metadata, schema only, or neither). In all conditions, participants were asked to provide their own interpretations; these text boxes aligned horizontally with the five dimensions in the Schema+Metadata and the Schema conditions.

Finally, after adding interpretations on all ten examples, participants reread the design brief and wrote their own proposed solution for the placemaking challenge (at least 50 and at most 500 characters). Participants were told not to perform additional web searches and that they would be disqualified if they copied a solution from the Internet.

Results

Our analysis focuses primarily on how a schema and metadata affects the interpretations offered for each placemaking example. We did not observe any differences between conditions in expert-rated novelty and usefulness on the final solutions for the placemaking challenge.

Schemas Yielded More Relevant Interpretations

To measure the relevance of participants' interpretations, we performed a semantic analysis of the relationship between generated interpretations and the example descriptions. After some simple preprocessing (i.e., removing stop words, spell correction, lemmatizing), we extracted 300-dimensional GloVe embeddings (Pennington, Socher, and Manning 2014) for each word in participants' interpretations and calculated a similarity metric based on the word vectors in the placemaking description. This analysis shows that both schema+metadata (b = 0.04, 95% CI: [0.01, 0.07], p = 0.005) and schemas alone (b = 0.05, 95% CI: [0.02, 0.08], $p < 0.001$) helped participants generate significantly more relevant interpretations than participants who just saw the examples with no schema (see Fig 2). There was no significant difference between schema+metadata and schemas alone in terms of relevance.

Schemas Alone Led to Less Novel Interpretations

While semantic similarity provides an indication of whether participants' interpretations are relevant versus off-topic, it does not tell us to what extent these interpretations overlap with language already in the placemaking example. Participants might just be copying text from the examples, where as a truly creative contribution should both stay on-topic *and* add something novel. Therefore, we analyzed how much participants simply copied words from the placemaking description, using measures of lexical similarity.

First, after removing stop words (e.g., "the"), we did a simple calculation of the number of overlap words between the interpretations and the example descriptions. We found that participants were significantly more likely to adopt language from the example description in schema-only condition, compared to schema+metadata (b = 1.27, 95% CI: [0.36, 2.17], p = 0.008) or compared to baseline (b = 1.51, 95% CI: [0.62, 2.41], p = 0.001) (see table 2).

However, the raw number of overlapping words between the intepretations and example descriptions does not account for total amount of words. Overlap is likely higher for longer example descriptions because there are more opportunities

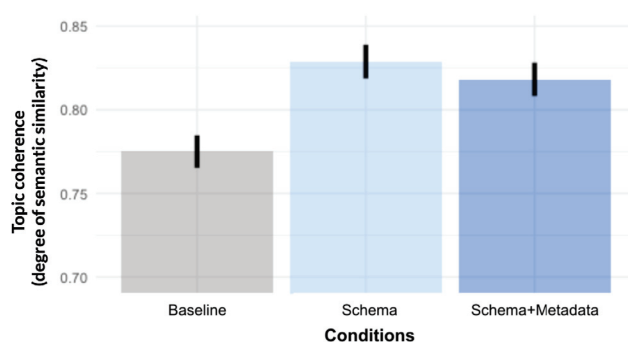

Figure 2: Both the schema and schema+metadata conditions yielded interpretations with better topic cohesion with the example text. Baseline participants were more likely to produce off-topic interpretations.

Condition	# of Words	# of Overlap Words	Jaccard Similarity
Baseline	2.40 (1.34)	0.55 (0.89)	0.023 (0.036)
Schema	2.61 (1.47)	0.93 (1.10)	0.039 (0.044)
Schema+ Metadata	3.08 (1.40)	0.65 (0.89)	0.026 (0.035)

Table 2: The average word count (and st. dev) in each interpretation, the average number of words that overlap with the example text, and the ratio of overlapping vs. non-overlapping words. Participants in the schema+metadata condition produced longer interpretations than the other conditions and also had a low ratio of overlapping words compared with participants in Schema condition. (Note: All word counts are calculated after removing stop words.)

for interpretations to overlap. To account for length, we calculated the Jaccard similarity coefficient:

$$\frac{num\ of\ overlapping\ words\ between\ interpretations\&example\ text}{total\ num\ of\ words\ in\ interpretations\ and\ example\ text}$$

The Jaccard similarity measure also shows a similar trend where the schema-only condition show a higher lexical similarity between the example description and the interpretations, compared to schema+metadata (b = 0.057, 95% CI: [0.02, 0.09], p = 0.002) and compared to baseline (b = 0.060, 95% CI: [0.02, 0.09], p = 0.001). Taken together, the number of overlapping words and the Jaccard similarity measure suggest that providing metadata has the benefit of diversifying the language in subsequent interpretations. Schema+metadata participants were less likely to simply copy surface details.

Discussion

Our experiment investigated how schemas and metadata affect cognition around crowd work. We found that expert-derived schema helped people to stay on task and on topic with their interpretation, while the metadata led to more diverse interpretations, but still within the schema constraints. These two strategies together create an effective knowledge framework for generating both relevant and diverse contributions on collective creativity project. Here we explore what drives these effects and how to leverage them in crowdsourcing.

Schemas Help Focus Attention on Key Details

Schemas, like good task instructions, can help crowd workers focus on what is most important when doing a task. Within the context of creative work, where the task is open-ended and difficult to assess with agreement-based quality control measures (Chilton et al. 2013), schemas provide helpful scaffolding (Vygotsky 1980). They provide a basic structure for how to think about a problem, and what to focus on, without being overly prescriptive.

While this preliminary study introduces just one type of schema, one can imagine a wide range of schemas, both within design where designers have been shown to decompose their thinking in multiple ways (Studer et al. 2018), as well as in other task domains (e.g., schemas for writing a request email(Hui, Gergle, and Gerber 2018)). Future work might seek to uncover the differences in how experts decompose their work or explore the effects of different schemas on novices' task performance.

Metadata Spark Divergent Thinking

When crowds perform creative tasks, the results are often mundane and/or repetitive (Chou and Tversky 2020), which is both inefficient and ineffective. Prior work on crowd creativity also shows that people recycle or reuse details from prior examples and recombine them in interesting ways (Yu and Nickerson 2011). Our study aligns with these insights to show how metadata can lead to more divergent contributions, essentially playing the role of inspiration during an ideation task. More importantly, we show how metadata can work in parallel with schema to yield divergent contributions within certain constraints. Prior work shows that creativity thrives within constraints (Costello and Keane 2000). The constraining effects of schema might, in fact, lead to better ideation than just providing metadata with no organization, although this would require additional study.

Effects Did Not Transfer to a Secondary Design Task

While we observe the dual effect of schema and metadata on the novelty and topic relevance of the interpretations, we did not measure any downstream effects on the design proposals generated by participants after viewing and interpreting all ten examples. In other words, adding novel and on-topic interpretations did not translate to the secondary task of creating a new solution. Perhaps participants took inspiration from the placemaking examples, regardless of whether they were exposed to interpretations by others. Future work could explore the impact of applying the schema and metadata strategy during the design proposal task.

Schemas and Metadata Can Impact Crowdsourcing

The strategies of providing schema and metadata can impact crowdsourcing platforms in multiple ways. Task designers can generate schema based on their own expertise and pull out early samples of crowd work to exemplify the range of diversity desired on a creative task. Schema generation could become part of the basic structure for crowd platforms, especially when an effective schema is not known a priori. For example, platforms might recruit successful workers to describe *how* they they perform a complex task, and then translate this into a schema that can be offered to subsequent workers. More research is needed to understand how different configurations of schema and metadata (e.g., number of schema attributes, amount and diversity of prior insights in metadata) affect thinking around a task.

Conclusion

This study shows how schema and metadata can improve performance on creative crowd tasks. Participants viewed placemaking solutions, with or without the presence of schema and metadata, and then offered their own creative interpretations of those examples. Schemas helped workers pay attention to key attributes, but when provided on their own also appeared to stymie creativity. However, when schemas are used in conjunction with diverse responses from others (i.e., metadata), participants made contributions that are both useful and novel. Future work could explore the impact of employing different schemas, automatically inferring schemas from crowd behavior, as well as novel approaches to distilling insights from other users to enhance creative work.

Acknowledgements

We thank the National Science Foundation for funding this work under grant #1122206.

References

André, P.; Kittur, A.; and Dow, S. P. 2014. Crowd synthesis: Extracting categories and clusters from complex data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 989–998.

Barber, S. T. 2018. The zooniverse is expanding: crowdsourced solutions to the hidden collections problem and the rise of the revolutionary cataloging interface. *Journal of Library Metadata* 18(2):85–111.

Brewer, W. F. 1987. Schemas versus mental models in human memory. *Modelling cognition* 187–197.

Chiang, C.-W.; Kasunic, A.; and Savage, S. 2018. Crowd coach: Peer coaching for crowd workers' skill growth. *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW):1–17.

Chilton, L. B.; Little, G.; Edge, D.; Weld, D. S.; and Landay, J. A. 2013. Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1999–2008.

Chilton, L. B.; Kim, J.; André, P.; Cordeiro, F.; Landay, J. A.; Weld, D. S.; Dow, S. P.; Miller, R. C.; and Zhang, H. 2014. Frenzy: collaborative data organization for creating conference sessions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1255–1264.

Chou, Y.-Y. J., and Tversky, B. 2020. Changing perspective: Building creative mindsets. *Cognitive Science* 44(4):e12820.

Costello, F. J., and Keane, M. T. 2000. Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science* 24(2):299–349.

Girotto, V.; Walker, E.; and Burleson, W. 2017. The effect of peripheral micro-tasks on crowd ideation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 1843–1854. New York, NY, USA: Association for Computing Machinery.

Goucher-Lambert, K., and Cagan, J. 2019. Crowdsourcing inspiration: Using crowd generated inspirational stimuli to support designer ideation. *Design Studies* 61:1–29.

Hui, J. S.; Gergle, D.; and Gerber, E. M. 2018. Introassist: A tool to support writing introductory help requests. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

Kittur, A.; Smus, B.; Khamkar, S.; and Kraut, R. E. 2011. Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 43–52.

Kittur, A.; Peters, A. M.; Diriye, A.; and Bove, M. 2014. Standing on the schemas of giants: socially augmented information foraging. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 999–1010.

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

Lee, D. J.-L.; Lo, J.; Kim, M.; and Paulos, E. 2016. Crowdclass: Designing classification-based citizen science learning modules. In *Fourth AAAI Conference on Human Computation and Crowdsourcing*.

Liu, X., and Bagrow, J. P. 2017. Autocompletion interfaces make crowd workers slower, but their use promotes response diversity. *arXiv preprint arXiv:1707.06939*.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Project for Public Spaces. 2012. Lighter, quicker, cheaper.

Schneekloth, L. H., and Shibley, R. G. 1995. *Placemaking: The art and practice of building communities*. Wiley New York.

Studer, J. A.; Daly, S. R.; McKilligan, S.; and Seifert, C. M. 2018. Evidence of problem exploration in creative designs. *AI EDAM* 32(4):415–430.

Vygotsky, L. S. 1980. *Mind in society: The development of higher psychological processes*. Harvard university press.

Yilmaz, S.; Seifert, C.; Daly, S. R.; and Gonzalez, R. 2016. Design heuristics in innovative products. *Journal of Mechanical Design* 138(7).

Yu, L., and Nickerson, J. V. 2011. Cooks or cobblers? crowd creativity through combination. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1393–1402.

Zhang, A. X.; Verou, L.; and Karger, D. 2017. Wikum: Bridging discussion forums and wikis using recursive summarization. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2082–2096.