Using Crowds to Study Creativity

Steven Dow

HCI Group
Department of Computer Science
Stanford University

spdow@stanford.edu

ABSTRACT

How does the creative process affect design results? Our research examines how prototyping practices affect learning, motivation, communication, and results in design. Our studies call on online participants to perform tasks where the solutions are creatively diverse *and* objectively measurable. Participants create Web banner advertisements. We then place the crowd-designed ads in an online campaign, collect a host of analytics (e.g., click-through data), and statistically compare performance differences. The key insight enabling our research is that crowdsourcing techniques and web analytics provide an opportunity to do experimental research on creativity with objective outcomes.

Categories and Subject Descriptors

H.1.m [**Information Systems**]: Models and Principles

General Terms

Design, Experimentation.

Keywords

Creativity, crowds, experimentation, design thinking, prototyping, advertising, Web analytics.

1. INTRODUCTION

Design shapes the world, both physical and digital. To produce better products, services and communities, we need a deeper understanding of how and why creative practices affect results. Our research seeks empirical evidence to strengthen our theoretical understanding of the creative process. What's our approach? We recruit crowds for creative work, collect analytic data on their designs, and then statistically compare performance differences. We use this experimental approach to manipulate and understand the effects of variables in the creative process.

Our crowdsourcing experiments engage participants in creative activities. Specifically, they design text-based or graphical Web advertisements for a real client. For example, a recent experiment generated hundreds of unique ad concepts for FaceAIDS.org, an organization dedicated to mobilizing students to fight AIDS in Africa. To measure design performance, we post all the ads online simultaneously and measure click-through rates and visitor analytics. We learn how hundreds of different creative ideas perform comparatively in an authentic advertising context. We also recruit stakeholders—such as the clients, other crowd workers, and ad professionals—to rate ad quality.

To enable this research, we built a web-based experimental platform that extends Amazon Mechanical Turk (AMT) in three ways. First, the system recruits and randomly assigns participants to experimental conditions. Second, our system supports real-time interaction with workers, allowing us to directly manipulate the critique process, where workers get feedback on their design ideas. This feature will also help us examine social interaction issues between crowd members. Third, after collecting design concepts (i.e., Web ads), the system automatically places results back into AMT where independent, blind-to-condition raters judge performance. We are currently developing the final step in this pipeline: automatically launching the online ad campaign to gather click-through analytics.

Expert critiques often improve design results and help novices learn to be better designers. Yet many crowdsourcing platforms lack effective feedback channels to present this type of critique. Our results show that by enabling real-time communication—and providing iterative critique—crowds produce better overall designs. Paradoxically, iterative critique can also give rise to fixation—continuously refining an idea without exploration. Another study found that creating prototypes and receiving feedback on multiple designs in parallel — as opposed to serially — leads to more divergent concepts, more explicit comparison, less investment in a single concept, and better overall design performance [1]. Structuring a more parallel design process will improve the results of crowdsourced creative work.

In addition to increasing our understanding of human creativity and design thinking, these experiments shed light on the potential for crowdsourcing creative work. As a whole, crowds may produce diverse concepts, but how often are the ideas original and economically viable? When can crowds compete with a few creative professionals? We are examining how particular attributes make creative tasks more suited for crowds. Finally, assuming the crowd produces at least a few gems, can the crowd also help pick them? We are currently developing statistical models to predict future design performance from crowd ratings.

2. ACKNOWLEDGMENTS

3. REFERENCES

[1] Dow, S., Glassco, A., Kass, J., Schwarz, M., Schwartz, D.L., and Klemmer, S.R. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy. Transactions on Computer-Human Interaction 4, (2010).