Advancing Large-Scale Creativity

through Adaptive Inspirations and Research in Context

by

Victor Augusto da Silva Girotto

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved December 2018 by the
Graduate Supervisory Committee:

Erin Walker, Co-Chair
Winslow Burleson, Co-Chair
Ross Maciejewski
Sharon Hsiao
Jeffrey Bigham

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

An old proverb claims that “two heads are better than one”. Crowdsourcing research and practice have taken this to heart, attempting to show that thousands of heads can be even better. This is not limited to leveraging a crowd’s knowledge, but also their creativity—the ability to generate something not only useful, but also novel. In practice, there are initiatives such as Free and Open Source Software communities developing innovative software. In research, the field of crowdsourced creativity, which attempts to design scalable support mechanisms, is blooming. However, both contexts still present many opportunities for advancement.

In this dissertation, I seek to advance both the knowledge of limitations in current technologies used in practice as well as the mechanisms that can be used for large-scale support. The overall research question I explore is: “How can we support large-scale creative collaboration in distributed online communities?” I first advance existing support techniques by evaluating the impact of active support in brainstorming performance. Furthermore, I leverage existing theoretical models of individual idea generation as well as recommender system techniques to design CrowdMuse, a novel adaptive large-scale idea generation system. CrowdMuse models users in order to adapt itself to each individual. I evaluate the system’s efficacy through two large-scale studies. I also advance knowledge of current large-scale practices by examining common communication channels under the lens of Creativity Support Tools, yielding a list of creativity bottlenecks brought about by the affordances of these channels. Finally, I connect both ends of this dissertation by deploying CrowdMuse in an Open Source online

community for two weeks. I evaluate their usage of the system as well as its perceived benefits and issues compared to traditional communication tools.

This dissertation makes the following contributions to the field of large-scale creativity: 1) the design and evaluation of a first-of-its-kind adaptive brainstorming system; 2) the evaluation of the effects of active inspirations compared to simple idea exposure; 3) the development and application of a set of creativity support design heuristics to uncover creativity bottlenecks; and 4) an exploration of large-scale brainstorming systems' usefulness to online communities.

ACKNOWLEDGMENTS

I am forever in debt to both of my advisors, Winslow Burleson and Erin Walker, for their guidance, expertise, and patience throughout this long academic journey. Win opened the doors for me. Without his support, I would not have stayed at ASU to pursue my doctorate. It was also through his lectures and multidisciplinary work that I learned about the exciting field of creativity. Erin kept me moving. Without her precise feedback, my research would have crumbled. Many of the findings described here were only possible because she correctly encouraged me to persist at the analysis. I am also grateful to my committee members, Ross Maciejewski, Sharon Hsiao, and Jeffrey Bigham.

I must also thank my colleagues at the 2 Sigma Learning lab, Nichola Lubold, Shang Wang, and Ishrat Ahmed, for kindly sharing their knowledge and time, as well as breaking my systems before study participants did. I must also thank Rachana Rao, Deniz Sonmez, Sameena Hossain, and John Frens for their help with data analysis and coding. Finally, I am grateful to Cecil Lozano and Kasia Muldner, who were strong positive influences during the early years of my doctorate.

I would not be writing these words if it were not for my wife Miriam, who gave me the love and encouragement to persist after every failed experiment. I am also thankful for the support received from my family—my mother Claudia and stepfather Ira, my father Guilherme and stepmother Noemi, my brother Gustavo, and grandparents Eser, Leila, Adamor, and especially Eurides, my academic role-model. I must also thank Jadiel de Armas for his friendship and advices that made this doctorate possible.

TABLE OF CONTENTS

Page

LIST OF TABLES ... vii

LIST OF FIGURES ... viii

CHAPTER

1 INTRODUCTION ... 1

Research Problem ... 6

Contributions ... 7

2 BACKGROUND WORK ... 12

Creativity ... 12

Large-Scale Creativity ... 28

Conclusion ... 36

3 BOTTLENECKS IN CURRENT COMMUNICATION CHANNELS ... 37

FOSS Communication Channels ... 39

CST Design Heuristics ... 43

Method ... 49

Results ... 52

Discussion ... 61

Limitations ... 63

Conclusion ... 65

4 IMPROVING INSPIRATION EFFECT THROUGH MICROTASKS ... 67

The System ... 69

Metrics ... 72

Experiment 1: Rating Task ... 75

Experiment 2a & 2b: Similarity Choice Task ... 78

Experiment 3: Comparison Across Task Types ... 85

Discussion ... 88

Conclusion ... 93

5 CROWDMUSE: SUPPORTING BRAINSTORMING THROUGH MODELING AND ADAPTATION ... 95

The CrowdMuse System ... 97

Study 1 ... 108

Study 2 ... 116

Revisiting the Categorization Scheme ... 119

Discussion ... 121

Limitations & Future Work ... 123

Conclusion ... 126

6 DEPLOYING CROWDMUSE IN AN ONLINE COMMUNITY ... 127

Deployment Community: FreeCAD ... 130

Method ... 131

Results ... 133

Discussion ... 144

Limitations ... 148

Conclusion ... 150

CHAPTER Page

7 FINAL DISCUSSION 151

Discussion..154

Limitations..165

Future Work 166

Conclusion..169

REFERENCES 170

LIST OF TABLES

Table Page

1. Peripheral Task Types and Examples 32

2. List of Foss Projects That Were Analyzed....................42

3. Distribution of Content Coding Across Communities and Heuristics 51

4. Relationships between Design Principle, Design Element, and Issues....................53

5. Fluency and Inspiration Metrics for Experiment 1 77

6. Breadth, Depth, and Influence for Experiment 1 77

7. Fluency and Inspiration Metrics for Experiment 2a 80

8. Breadth, Depth, and Influence for Experiment 2a 80

9. Fluency and Inspiration Metrics for Experiment 2b 82

10. Breadth, Depth, and Influence for Experiment 2b....................82

11. Fluency and Inspiration Metrics for Experiment 3....................85

12. Breadth, Depth, and Influence for Experiment 3....................86

13. Marginal Means (and Standard Error) for Breadth....................114

14. Fluency and Inspiration Metrics for Study 2 118

15. Usage Statistics Across Chapter 5 and 6 Studies....................135

16. Perceived Benefits and Issues and Their Relation to CST Heuristics....................146

LIST OF FIGURES

Figure Page

1. Screenshot of the Ideation System as Used in the Final Experiment....................71

2. Snapshots of the Ideation Tree and Idea Pool in Five Different Points of Time .74

3. Rating Task Interface...76

4. The Task Panel for Condition 3 79

5. Regression Lines for Breadth by Fluency 87

6. Marginal Means and Std. Error for Breadth on the Second Half of Ideation 88

7. The Crowdmuse System...97

8. Crowdmuse Popups 99

9. Interaction Between Adaptive Solution Space and Inspirations on Fluency......113

10. Final Configuration of the Solution Space 136

CHAPTER 1

INTRODUCTION

Creativity is arguably one of the most important phenomena that can be studied, as it leads to novel and useful advances. Throughout the years, the way this phenomenon has been understood has undergone dramatic shifts, transitioning from mythical muses that are beyond the realm of scientific scrutiny (Sternberg & Lubart, 1999) to deep and insightful investigations of individuals and their traits (Guilford, Merrifield, & Wilson, 1958) and, more recently, groups and social settings (Osborn, 1963). This latter transition is particularly fitting in times where problems are increasingly multi-disciplinary, and innovation frequently emerges from groups and communities rather than exclusively from the insights of a few gifted minds.

Given the advances in communication technologies, such groups have been undergoing transformations, particularly in their scale and, consequently, their diversity. This holds great potential for creativity (Fischer, 2005). In practice, this increasing scale is materialized in open innovation initiatives such as InnoCentive[1] and OpenIDEO[2]—platforms that host innovation challenges open to anyone willing to solve them—as well as specific initiatives such as Netflix's open challenge[3] to increase their accuracy in recommending items.

But why is large-scale creativity so appealing? The answer can be estimated from the benefits found in smaller groups. One of the main process gains in group ideation is

[1] https://www.innocentive.com
[2] https://openideo.com
[3] https://netflixprize.com/

synergy, that is, one person may build on ideas proposed by others (Dennis & Williams, 2003). These synergistic ideas would hardly be elicited if those individuals worked in isolation. At a larger scale, the sheer volume of generated ideas may improve the chances of someone seeing another idea that sparks that synergistic insight. Groups can also have positive effects in idea selection due to their information asymmetry. This means that different people will have different parts of the information necessary for making an informed decision (Stasser, 1992), such as choosing the best idea for implementation. Therefore, groups should generate more ideas and could make more informed decisions due to the increased scale and diversity of the group. These and other advantages have the potential to be even greater in crowds. In practice, innovative solutions forged in response to the increasing number of innovation challenges pose as further evidence to the benefits of creativity at scale.

Opportunities to see this large-scale creativity flourishing abound throughout the internet, such as in the aforementioned InnoCentive and OpenIDEO initiatives. But arguably one of the best examples of crowd-scale creativity exists in Free and Open Source Software (FOSS) communities. These communities have produced remarkable and well-known results, such as the Apache server, the Linux operating system, and a plethora of other tools widely used by many—all of this despite most contributors not being geographically co-located and operating on their free time. This is only made possible by the many communication tools employed by these communities, such as discussion boards, chat systems, and versioning control systems. For many of these projects, it is in great part through these channels that new ideas are discussed and creative new directions for the projects can be undertaken. However, these tools were designed before we knew more about how to design tools meant to support creativity (such as the conclusions of the 2005 NSF workshop on Creativity Support Tools, as reported by Shneiderman et al., 2006), and therefore may fail to adequately support innovation within these communities, at least to its full potential.

Parallel to these issues, research has been exploring how to use large crowds for creative endeavors. This means dealing with problems brought about by the same scale that they hope would boost ideation. These issues are unique to this recent large-scale context, and therefore can't find clear answers in the creativity literature. For example, the sheer amount of ideas generated can hinder the synergistic performance, since an individual is unlikely to be able to read all of the ideas (thus possibly missing the one that could inspire them), much less pay attention to them, which is a necessary requirement for influence (Nijstad & Stroebe, 2006). To address these and other issues, research has been investigating how to improve large-scale creativity, with a special emphasis on brainstorming support. Support usually comes by showing ideators some form of **inspiration**—usually others' ideas, or something created by a crowd or facilitator. These inspirations are many times created by Amazon Mechanical Turk workers through microtasks—small tasks that can be done quickly and with low effort. For example, it has been found that brainstorming can be improved by having experts facilitating the ideation (Chan, Dang, & Dow, 2016b), showing ideators a diverse set of ideas (Siangliulue, Arnold, Gajos, & Dow, 2015), carefully designing the timing of inspiration delivery (Siangliulue, Chan, Gajos, & Dow, 2015), by leveraging analogies (Yu, Kittur, & Kraut, 2014b, 2014a; Yu, Kraut, & Kittur, 2016), or by having computers abstract others' ideas and using that as inspirations (Chan, Dang, & Dow, 2016a). The common thread among this research is that of using humans or computers (or both) to generate some form of inspiration (such as an analogy, a diverse set of ideas, a text suggestion) that can be shown to ideators to improve their performance. The focus, therefore, is on the inspirations themselves rather than on the ideators they're supposed to inspire.

There are two ways for improving these support approaches: increasing their **effectiveness** and **relevance** to ideators. Improving effectiveness means increasing the effect an inspiration may have on the ideator that sees it. Most forms of inspiration investigated so far are passive, simply asking users to read something—meaning they might not tend enough to the inspiration they are exposed to. This is a limitation, as it has been found that attention is key for ideators to be influenced (Brown, Tumeo, Larey, & Paulus, 1998; Coskun, Paulus, Brown, & Sherwood, 2000; Nijstad & Stroebe, 2006). This is somewhat parallel to the discussion around active vs. passive learning in educational research (Chi & Wylie, 2014; Prince, 2004), generally showing an advantage towards active methods. Therefore, exploring ways of increasing the attention ideators pay to inspirations could bring positive implications to current approaches of ideation support.

Improving relevance, on the other hand, means ensuring that you are showing ideators something that they are likely to be inspired by. Currently, inspirations are generally chosen at random (e.g. Chan et al., 2016a) or based on their intrinsic qualities or diversity (e.g. Siangliulue, Arnold, et al., 2015). However, important creativity models show that ideators have specific cognitive structures and may be more likely to be influenced if exposed to certain categories of ideas than others (Brown et al., 1998; Charlan Jeanne Nemeth & Nemeth-Brown, 2003; Nijstad & Stroebe, 2006).

Consequently, an inspiration selection mechanism that doesn't take individual ideators into consideration may be missing the opportunity to appropriately inspire them. Therefore, there is an opportunity for exploring ways of modelling ideators in real time in order to inform inspiration selection.

Both improvements—to effectiveness and relevance—relate to improvements to current techniques explored in research. However, on a broader view, there are other yet unexplored avenues in the literature. First, this research has mostly been done through paid crowd labor markets such as Amazon's Mechanical Turk (MTurk), meaning that ideators were likely involved in the task for the financial compensation—a form of motivation that has been found to not be conducive to creativity (T. M. Amabile, 1985). In addition, ideation usually happens for only a few minutes, while it has been suggested that stronger results in creativity support would only emerge from studies of longer duration (Shneiderman et al., 2006). Furthermore, problems have been mostly simple and domain general, such as designing a chair (Yu & Nickerson, 2011) or writing a birthday card message (Siangliulue, Arnold, et al., 2015). While most people could have some knowledge to contribute with ideas towards those problems, one could arguably claim that people may not necessarily be motivated to do so (or, relating to my last point, may be motivated only by the financial gains). This would negatively affect the expected creative output (T. Amabile, 1983).

None of these limitations deny the validity of the contributions made by this line of research. Nonetheless, the literature suggests that these factors could be interfering with the results, or potentially reducing their applicability. Furthermore, many have approached this research by building systems and evaluating them in crowd markets.

Examples include IdeaHound (Siangliulue, Chan, Dow, & Gajos, 2016), IdeaGens (Chan et al., 2016b), BlueSky (Huang & Quinn, 2017), and CrowdMuse (Chapters 5 and 6 of this dissertation). But would these systems be useful outside of crowd markets? For example, could these systems be adopted as part of the workflow in existing online communities, replacing some of the outdated tools that are currently used? Answering these questions is key to ensuring that the exciting developments made by the crowdsourced creativity researchers are useful and applicable to current real-world creative communities.

Research Problem

Considering the issues and possibilities, in this dissertation I explore the following overarching research problem: "**How can we support large-scale creative collaborations in distributed online communities?**" I approach this overarching question both by looking into current practices in online communities, as well as advancing the current crowd creativity research agenda. Namely, I deal with the following research sub-questions:

1) *What are the creativity bottlenecks brought by the communication channels currently employed by large-scale distributed communities?* The communication channels employed by online communities are usually traditional forms of communication such as discussion boards or chat interfaces, which are designed for discussion. These tools do not take advantage of recent advances in creativity support design, and when used for

such purposes, may hinder creativity. However, it is unclear what are the issues, and what causes them.

2) *How can we improve attention to inspiration interventions?* Current approaches for improving idea generation are generally passive, meaning that they simply show ideators a snippet of text. However, attention is a key process behind inspiration, and therefore it is desirable to increase it.

3) *How can we improve the relevance of inspirations?* Current inspiration approaches have focused on evaluating differences in inspiration strategies but are yet to focus on differences with ideators themselves. Idea generation models posit that different people have different approaches to ideation, and therefore are likely to respond better to some kinds of inspirations than to others.

4) *How could a crowd brainstorming system be integrated into an existing online community?* The advances made by crowd creativity research often result in systems that could be useful for online communities. However, it is unclear whether they would appropriately meet these communities' needs. Therefore, exploring this question can provide guidance for future crowd creativity systems to ensure they may be relevant and useful in real-world scenarios.

Contributions

By exploring the research questions proposed above, this dissertation makes the following contributions:

- A list of creativity bottlenecks in current communication channels used by distributed online communities, as well as system UI affordances that may be leading to such issues;
- An understanding of the effects that performing microtasks as a way of increasing attention to inspirations can have on ideators;
- An approach for modeling and adapting to individual ideators, implemented through the CrowdMuse system, as well as an exploration of this approach's effects on ideation performance;
- A description of the deployment of CrowdMuse in an open source community, describing how the system was used, as well as the benefits and limitations users saw in it.

The remainder of this dissertation is structured as follows. Chapter 2 will discuss the relevant background work, starting with a discussion of creativity and idea generation. The purpose is to establish what creativity is, how it relates to idea generation, the underlying cognitive processes of idea generation, and how they can be supported. The latter will be exemplified mainly through a discussion of Creativity Support Tools (CST), and how they have tried to support creativity and idea generation. However, most of past research on creativity focused on individuals or small groups, though there have been recent pushes towards an increasingly larger scale. Therefore, the chapter will conclude with a discussion of crowdsourcing and large-scale creativity. The purpose is to outline how current large-scale creativity attempts look like, both in research and practice. This discussion will highlight three limitations: 1) the limitations of popular communication channels for idea generation; 2) limitations with current microtask-based research approaches; and 3) lack of exploration of the individuality of creativity. The following chapters will approach each of these three points in more detail.

Chapter 3 will take a closer look into popular communication channels used by distributed communities for creative collaboration. These channels include tools such as discussion boards, mailing lists, and bug trackers. This analysis will be done through the lenses of CSTs and their many design principles that have been developed over time. The question asked is: if distributed communities engage in creative collaborations through these tools, and these tools were not designed with creativity in mind, what could be their negative effects on the communities' creativity? Using Free and Open Source Software (FOSS) communities as a case study, I explore this question through heuristic and content analyses, as well as interviews with community members. The outcome is a set of limitations and the affordances in the channels' interfaces that could be causing them.

In Chapter 4, I temporarily part from discussing crowd creativity in practice, and turn to addressing limitations in current research approaches. As will have been established in the background work (Chapter 2), much of crowdsourced creativity research relies on support generated by external crowds—generally those from microtask markets such as MTurk. For example, these crowds can rate the ideas, with the best being used as inspiration for the ideation crowd. This poses a clear limitation when the task at hand demands certain privacy or extensive knowledge of the domain. This chapter, therefore, explores the usage of the work usually offloaded to a third-party crowd as a form of inspiration to the crowd itself. The work presented in this chapter is based on Girotto, Walker, & Burleson (2017).

Having considered issues in both research and practice, Chapter 5 begins to explore a new avenue for supporting crowd idea generation. While current research has explored supported based on content—that is, analyzing differences in inspiration types—this chapter deals with support based on the individual—that is, adapting the inspiration based on the traits of the ideators themselves. To this end, this chapter introduces CrowdMuse (Girotto, Walker, & Burleson, 2018), a system which models ideators based on their past ideas' categories. The model keeps track of the categories visited, transitions between categories, frequency of category changes, and other attributes. By leveraging recommender system techniques together with these models, the system is also able to suggest new categories to individual users. All of this translates into two types of adaptation: *subtle*, in which the views of the system are reordered based on the model; and *explicit*, in which the inspirations presented on-demand to users are chosen based on the model. The system is evaluated through two studies on a crowd platform, from which I draw implications related to the different types of adaptations as well as the effect of the level of the categories used to power them.

In chapter 6, I once again approach the theme of large-scale creativity in practice, in order to connect it with the research described in the previous two chapters. The intent is to explore the usefulness of crowd brainstorming systems such as CrowdMuse in a real-world context. To achieve this, I deploy CrowdMuse on an open source community for two weeks. Using system logs, posts in the discussion thread, and participant interviews, I describe how users behaved in the system, contrasting it to the behavior observed in the crowd studies in Chapter 5. I also uncover a list of benefits the system brought, as well as issues it did not address. I finalize the chapter with a discussion of the implication these findings can have on the design of crowd creativity support systems.

Finally, in Chapter 7 I conclude with an overall discussion of the connection between the contributions made in each chapter, as well as their implication for the advancement of crowdsourced idea generation research and practice.

CHAPTER 2

BACKGROUND WORK

Creativity

Perspectives on Creativity

Among the many human traits valued by a society that incessantly seeks innovation, creativity arguably ranks among the top. While earlier perceptions framed creativity under a mysterious, if not mythical light, only relatively recently has research started to explore this topic methodically (Guilford, 1950). Inspired by historical and contemporary geniuses, inventors, and artists, research initially focused on the individual. Subsequently, the role that groups and organizations play in this process has been increasingly acknowledge and explored, especially in face of problems that require interdisciplinary creativity.

What is creativity? Simply put, it can be defined as the generation of something new and useful (Hennessey & Amabile, 2010). However, this simple definition only scratches the surface of the richness of the domain. There are multiple views on the *focus of analysis* (what aspect of creativity is being studied), *levels of magnitude* (how impactful the creative outcome is), and *orientations* (the perspective taken to study creativity) (Kozbelt, Beghetto, & Runco, 2010).

Focus of Analysis. The initial focus for modern creativity research was on the individual, with J.P. Guilford (Guilford, 1950), who focused on the measurement of individuals through tests—the so called psychometric approach (Sternberg & Lubart, 1999). According to this approach, creativity was usually examined by applying divergent thinking tests such as Guilford's Unusual Uses Test (Guilford et al., 1958) and the Torrance Tests of Creative Thinking (Torrance, Ball, & Safter, 2003), which listed a simple proposition (e.g. list unusual uses of a cardboard box) and examined, for example, the number of different responses elicited (Sternberg & Lubart, 1999). While this approach concerned itself with measuring individuals through the artifacts they developed (i.e. the answers to the tests), there are other points of focus. These can be summarized in the six P's of creativity: process, product, person, place, persuasion, and potential (Kozbelt et al., 2010; Runco, 2014). For example, an approach focusing on product will examine artifacts such as paintings and test answers as they relate to creativity, while a process-focused approach will consider the process or stages leading to a creative result.

The approaches based on the creative process are of particular interest in this dissertation. At the broadest level, one can distinguish between divergent and convergent phases of thought (Cropley, 2006; Kozbelt et al., 2010). The divergent phase is characterized by the generation of diversity, variation, multiplicity of choices. A prime example of this kind of thought happens in brainstorming sessions, where the goal is to generate as many ideas as possible—the more original the better. Having done that, people then move on to organizing, evaluating, tweaking, and selecting ideas. This is the convergent phase, where the goal is to rationally evaluate ideas to select the best one. While divergent thinking has usually been the focal point of interest (e.g. the previously mentioned divergent thinking tests), convergence is equally as important. Divergence without convergence is likely to lead to disastrous or ineffective changes (Cropley, 2006).

The divergent/convergent dichotomy provides a useful overall distinction of phases in the creative process, but finer grained models exist. Among the earliest is Wallas' four-phase model (Wallas, 1926, as reported in Kozbelt et al., 2010), where an individual goes through the stages of *preparation* (gathering knowledge), *incubation* (time away from the problem, unconscious work), *illumination* (the solution appears in a classic "aha" moment), and *verification* (validation of the solution). More recently, Cropley & Cropley (2008) essentially extended this model to include 7 phases. The preparation phase is now followed by the activation phase, in which problem awareness is developed. Incubation turns into cogitation, making this phase more flexible than incubation (e.g. it can now contain processes such as ideation). More importantly, the verification phase is followed by two phases: communication—the product is revealed to significant stakeholders, and validation—the product is validated by those stakeholders. This extended model, therefore, agrees with Csikszentmihalyi's system perspective (Csikszentmihalyi, 1999), describe later in this chapter. However. the linear nature of both models fails to capture the recursive, iterative, and perhaps relatively chaotic reality of the process. Cropley & Cropley acknowledge this limitation, noting however that a model such as this highlights the processual nature of creative (as opposed to a simple event), and allows for a qualitative difference between phases (that is, different strategies and incentives may be beneficial at different points of this process). This dissertation focuses on the phases of cogitation (through idea generation and development) and illumination (as a consequence of supporting cogitation) phases.

Levels of Magnitude. A second way of differentiating approaches concerns the magnitude of the impact that creativity can have. Traditionally this differentiation has been conveyed through the four C's of creativity, as described by Kaufman & Beghetto (2009): *mini-c* creativity operates at a personal level, exemplified by insights that happen during learning experiences. A step above it, *little-c* creativity describes simple, everyday creativity, which can be experienced daily in simple things such as cooking new dishes or creatively solving domestic issues. *Pro-c* creativity takes this a step further, focusing on creativity exhibited by those whose creativity signifies professional level proficiency in a field. A good example perhaps is exhibited by researchers who consistently employ their creativity to advance their fields of study. However, while most researchers contribute to the advancement of a field, some can revolutionize it. To those, their creativity can be classified under *Big-C* creativity. This level is concerned with eminent creativity, with examples in the ranks of Jobs and Bach, and is exemplified in research that examines creativity through a biographical approach (e.g. Csikszentmihalyi (1997). While this level of creativity is the one commonly thought of when creativity is in question, it is important to be mindful of the other levels, which can have personal, social, or professional impact. The present work focuses on supporting anything between little-c and Pro-c.

Orientation. Finally, researchers have also examined creativity through diverse orientations, such as cognitive, evolutionary, and developmental (Kozbelt et al., 2010; Sternberg & Lubart, 1999). For this work, I highlight two perspectives, one at an individual level (componential) and another at a societal level (systems). On the individual level, Amabile (1983) developed the componential model of creativity, which defines creativity as being fueled by different components, which when maximized would increase the output of work deemed creative (per the consensual assessment definition, explained later in this chapter). Initially, she discussed three components: 1) domain-relevant skills, which comprise knowledge about and skills within a domain; 2) creativity-relevant skills, pertaining to strategies and styles for being creative; and 3) task motivation, that is, the individual's motivation towards the task. Therefore, if an individual shows high knowledge, creativity skills, and intrinsic motivation, he or she will likely be able to output creative products. More recently, the social environment was also included as a standalone component (T. Amabile, 2012; T. Amabile & Mueller, 2002). This component, as opposed to the previous three, is external, and can impose limitations on the individual's creativity. For example, a social context in which divergence is not well received can limit creative potential, even if the individual's other three components are well-developed. This has practical implications: increasing knowledge, incentivizing useful creativity strategies, and fostering motivation, all while keeping external pressures at a minimum, can all increase creativity. In this way, an educational tool can support creativity just as much as a system dedicated to idea generation.

On a societal level, Csikszentmihalyi (1997, 1999) describes a systems approach to creativity. According to this view, creativity occurs when an idea or product developed by one *individual* is selected by a *field* (a group of domain experts) to be included into the *domain* (the body of knowledge in an area of study). In a practical example proposed by Csikszentmihalyi (1999), Einstein (the individual) generated the formula for relativity. This formula was appreciated by several influential people in the domain such as university professors (the field), who acted as gatekeepers to the domain. Therefore, Einstein's ideas were eventually integrated into the domain body. A systems perspective implies that creativity cannot happen in an isolated context—it is social. Furthermore, the integration into a domain by a group of gatekeepers implies that this view on creativity happens at least at a Pro-c creativity level, as anything lesser likely will not be judged creative enough by the gatekeepers to be added into the domain.

Idea Generation

Perhaps the most popular imagery associated with creativity is that of a lightbulb lighting up on top of someone's head as a new idea suddenly manifests itself, or a group plowing through post-its during brainstorming sessions. These images relate to the process of idea generation or ideation, through which ideas are generated towards achieving a goal (e.g. solving a problem). Idea generation (or, as a generalization, divergent thinking) has been the main focus of much research, in lieu of other processes or phases of creativity (Cropley, 2006; Gabriel, Monticolo, Camargo, & Bourgault, 2016; K. Wang & Nickerson, 2017). Nonetheless, this divergent process is fundamentally necessary (although not sufficient) for creativity, so understanding how it works and how it could be supported can bring about great rewards. This process does not happen ex-nihilo, as if brought about by a mysterious muse. A fundamental requirement, as already discussed, is that the individual has knowledge in the domain they are ideating in (T. Amabile, 1983; Cropley & Cropley, 2008; Wallas, 1926).

Idea generation commonly happens in groups, employing methods such as Brainstorming. This method was proposed by Alex Osborn, and proposes a few simple rules for idea generation, such as holding back on criticism and building on the ideas of others (Osborn, 1963). Group brainstorming can be particularly advantageous since individuals may be prompted, upon hearing an idea suggested by someone else, to explore concepts that he or she would not have thought of otherwise. It also has the opportunity of leveraging diverse backgrounds and expertise, especially for interdisciplinary problems. However, these benefits do not always materialize. Most notably, it has been consistently found that brainstorming in groups (as opposed to individually) decreases performance (Dennis & Williams, 2003; Diehl & Stroebe, 1987, 1991), due to reasons such as free riding, evaluation apprehension and, mainly, production blocking (i.e. not being able to think about new ideas due to having to listen to someone else propose their own ideas) (Diehl & Stroebe, 1987, 1991). Some of these issues are due to the communication medium used in traditional brainstorming and can thus be fixed by switching it. For example, electronic brainstorming allows for concurrent idea generation, fixing production blocking (Dennis & Williams, 2003). Other issues, especially at convergent stages, are not so simple to resolve. For example, groups tend to work towards an artificial consensus through behaviors such as groupthink, majority influence, and polarization (Charlan Jeanne Nemeth & Nemeth-Brown, 2003), and fixation on ideas proposed by others (Jansson & Smith, 1991; Smith, 2003). These are challenging issues, requiring careful thought if they are to be successfully overcome.

Theoretical models can provide a deeper and clearer reasoning for the benefits and drawbacks involved in group ideation. These models are also key to the research presented in this dissertation, especially for designing the CrowdMuse system, described in Chapter 5. The first is the Search for Ideas In Associative Memory (SIAM) model (Nijstad, Diehl, & Stroebe, 2003; Nijstad & Stroebe, 2006), which describes idea generation in terms of memory recall. The assumption is that there are two memory systems: working memory (WM) and long-term memory (LTM). LTM is essentially unlimited, and is organized in images, which are central concepts (e.g. a computer) along with associated features (e.g. has a CPU, has a storage unit). WM is where conscious processing takes place but is quite limited. An additional component worth noting is the search cue, existing in the WM, which serves as the cue to search the LTM. This search cue is comprised of items such as the problem definition, previous ideas, or personal experiences. With these components in place, idea generation is then described in terms of two loops between the two memory systems. The first loop is the *image retrieval loop*, where an image is retrieved from LTM and loaded into the WM. At this point, the image and associated features are available as the basis for the second loop, the *idea production loop*. The individual can now produce ideas using the image, its features, and whatever is on the search cue. This goes on until no more ideas can be thought of using the current image. At this point, the individual reverts back to the first loop, searching for another image to be loaded into the WM. This process follows until no more images can be retrieved, ending idea generation. Operations involving LTM are slower than those in the WM, meaning that the first loop takes longer than the second. This model is extremely informative, allowing us to predict a few things such as the generation of ideas within a category being faster than those between different categories, trying to remember an idea will prevent new idea generation (as it occupies the WM), the user quitting after successfully failing to retrieve new images, and external influences only affecting performance if sufficiently attended to (thus being added into the search cue). This notion of attention being necessary for external influences to affect idea generation is extremely important, and is present in much of the idea generation literature (Brown et al., 1998; Coskun et al., 2000; Dugosh, Paulus, Roland, & Yang, 2000; Paulus & Brown, 2003).

The second model is the matrix model proposed by Brown, Tumeo, Larey, & Paulus (1998), which shares some similarities with SIAM. Operating at the level of categories of ideas, it follows the notion that there are categories that are more or less likely to be activated (that is, ideas generated within a given category). They represent this through a matrix of category transition probabilities. Rows and columns are the same, representing different categories, and each cell contains a number between 0 and 1, representing the probability of transitioning from a category to another. The diagonal of the matrix, therefore, represents the probability of staying within the same category (similar to SIAM's idea production loop), and the other cells represent the probability of transitioning to a different category (similar to SIAM's image retrieval loop). While this model does not explain the cognition underlying idea generation process in such detail as SIAM does (Nijstad & Stroebe, 2006), its quantitative nature affords computation, such as simulation and real-time modelling of users—although it is unclear how to derive the matrix of probabilities for individuals. It also allows for a numerical representation of different ideation styles, such as those who favor divergence (identified by higher between category transition probabilities) or those who favor convergence (identified by higher within category transition probabilities). Both models can be quite informative in research revolving around idea generation and contribute greatly to this research.

Idea Combination and Iteration

The ideas generated through initial ideation, even if creative, are likely to be immature or incomplete. Therefore, they could benefit from the developmental processes of combination (mainly a divergent process) and iteration (mainly convergent). Idea combination is not uncommon, as building on the ideas of others is one of the “rules of brainstorming (Osborn, 1963), and is at the center of one of the most important process gains of group ideation: synergy (Dennis & Williams, 2003). Furthermore, lab studies lend support to its importance. Idea combinations with common ideas can yield more impactful ideas, and in greater numbers, while combinations with uncommon ideas can yield more novel and feasible ideas (Kohn, Paulus, & Choi, 2011). Positive effects are especially evident when combinations happen between dissimilar concepts (Dahl & Moreau, 2002; Doboli, Umbarkar, Subramanian, & Doboli, 2014). This effect is in great part due to emergent properties (Chan & Schunn, 2015; Ward & Kolomyts, 2010), which are properties belonging to the concepts that are not salient in the concepts themselves, but when the combination occurs, they are made evident. For example, a spider may have the property “spins web” and “catches prey”, and a “human” may have the property “lives in a city” and values a functioning “societal structure”. A “spider-human” combination could have multiple emergent properties, one of them being “fight crime”—a property that would not be evident in any of the concepts considered individually.

While idea combination can positively contribute to creativity, it may not be sufficient. Considering combinations as a divergent process, they would also need a convergent counterpoint to ensure usefulness (Cropley, 2006). And even though divergence is commonly heralded as the pathway to creativity, some consider a more focused, convergent thinking to also be another pathway for creative results (Nijstad, De Dreu, Rietzschel, & Baas, 2010). This convergent counterpoint can come through the iterative refinement, or development, of ideas. Iteration is, therefore, an important convergent component for creativity. It can, for example, help those with no experience to perform as well as those with experience in some design tasks (S. P. Dow, Heddleston, & Klemmer (2009). But this fact is perhaps best visualized in Chan & Schunn's (2015) exploration of a large dataset of crowd-generated ideas. They have found that distant combinations are important, but not sufficient for achieving good results. Iteration is necessary to build maturity in ideas—later combinations were more useful than earlier ones. A similar effect was also found by Yu & Nickerson (2011). In an exploration of a human-powered genetic algorithm, they have found later designs, developed through successive combination and evaluation rounds, to be more creative than earlier ones. Consequently, it is critical to allow ideas to be developed further from their initial forms.

Evaluation of Creativity

Having examined different aspects of creativity, in particular idea generation, combination, and development, the question remains as to how to measure creativity. This is a particularly challenging proposition, given the apparent subjectivity of creativity—while people can perceive something as creative, it is hard to define why. Nonetheless, one of the main forms of creativity assessment takes advantage of this notion. The consensual assessment technique (T. Amabile, 1983; Dollinger & Shafran, 2005; Hennessey, Amabile, & Mueller, 2011) proposes that something is creative to the extent that appropriate evaluators agree it is creative. Therefore, a measurement of creativity can be derived for individual products or ideas within a set of items (e.g. the ideas generated in a brainstorming session) by asking people with considerable knowledge of a domain to evaluate their creativity relative to each other. This is quite reminiscent of the systems perspective (although in a smaller scale), in which creativity exists through the judgement of the gatekeepers—domain experts. This evaluation is done independently and in random order, without any form of definition of creativity being presented to the judges. Based on these independent evaluations, an inter-judge reliability measurement is then calculated. If a considerable level of reliability is found (e.g. above 0.7), the final creativity measurement can be extracted for individual items by summing or averaging their ratings. This method has seen successful application in a broad array of domains (Hennessey et al., 2011).

There are, however, other constructs for assessing creativity. As previously mentioned, initial research on creativity focused on evaluating individual creativity through tests of divergent thinking (Guilford et al., 1958; Torrance et al., 2003). These tests present objects to individuals and ask them to come with up with as many different uses for those objects as possible. The number of ideas generated based on those prompts serves as a general indicator of the creativity of individuals. Nevertheless, models like Amabile's componential model and SIAM propose that the domain of idea generation would have an influence on these results. Therefore, many metrics for evaluating ideation sessions exist (Plucker & Makel, 2010; Shah, Smith, & Vargas-Hernandez, 2003; Sternberg & Lubart, 1999), including:

- *Fluency* represents the quantity of ideas generated.
- *Elaboration* represents the average level of detail in each idea developed by an individual or group.
- *Flexibility/Divergence* quantifies the number of different categories represented by the ideas generated.
- *Convergence* represents the number of ideas within each category

- *Originality* is the statistical rarity of the answers
- *Usefulness* or feasibility represents the perception of how useful an idea is. Following the definition of creativity as something that is both original and useful (Hennessey & Amabile, 2010), this metric has been used in conjunction with originality to derive a measure of creativity outside of the consensual assessment (e.g. Chan et al., 2016b).

Scalable techniques for assessing creativity, generally derived from the metrics above, will be discussed later in this chapter.

Creativity Support Tools

The review above provides a snapshot of the current understanding of creativity. With this knowledge, a natural follow-up area of inquiry is whether we can design systems and tools to enhance creative performance. A promising evidence is seen in research that employs electronic brainstorming (EBS), which is able to improve performance just by addressing the issue of production blocking (Dennis & Williams, 2003). Production blocking happens when users stop generating ideas to pay attention to another group member's idea. EBS allows users to delay this attention to when its most convenient to them by keeping a record of the suggested ideas. Seeking even further creative performance gains, research has worked towards the development of Creativity Support Tools (CST). CST attempt to support creativity in its different phases, though most tools tend to focus on idea generation and divergent thinking (Gabriel et al., 2016; K. Wang & Nickerson, 2017).

Just as creativity is multifaceted, so are the ways of supporting it. Nakakoji (2006) categorizes CSTs in terms of their objective: they can act as dumbbells, building your creative strength over time; they can act as running shoes, which improve creative performance while the tool is used, but not over time; and finally, they can act as skis, which enable completely new experiences that relate in some way to a creative output. This distinction is important, Nakakoji argues, since they affect the evaluation of such tools. For example, dumbbell tools should be evaluated by “creative growth” over time, while running shoe tools should be evaluated by performance in the moment. The evaluation of ski tools is not as clear, although I would argue that this category is not mutually exclusive with the other two. While skis do allow for new modalities of support, it can still provide performance enhancement over time (e.g. the usage of skis can build muscle on your legs) or in the moment (if your goal is to get from point A to point B, skis may enhance your performance more than running shoes). Applying this to CSTs, it’s possible that “ski” tools can help you build your creative muscles, or simply enhance creative interactions in the moment, thus allowing ski tools to be evaluated similarly to the other two.

Lubart (2005) proposes a different classification, characterizing support tools in regards to their role during the creative process: they can act as nannies, monitoring and supporting emotional and cognitive aspects of the user; they can act as pen pals, supporting creative collaborations; they can act as coaches, embedded with expertise in creative processes and guiding users accordingly; and they can act as a colleague, leveraging advances in Artificial Intelligence to develop autonomy in creating or modifying artifacts in creative manners. Again, these categories are not mutually exclusive, as a tool can, for example, act bot has a coach and as a colleague, not only guiding the user in creativity relevant strategies, but also applying them itself.

Much of the research on CSTs revolves around the development or application of design principles. Many different sets of principles have been developed, and there is considerable overlap between them (I describe these principles in much more detail in Chapter 3). One meaningful contribution towards the development of standardized principles happened at the 2005 NSF-sponsored workshop on CSTs, where one of the outcomes was a list of 12 principles for designing CSTs (Resnick et al., 2005; Shneiderman et al., 2006):

1. Support exploration
2. Low threshold, high ceiling, and wide walls
3. Support many paths and many styles
4. Support collaboration
5. Support open interchange
6. Make it as simple as possible—and maybe even simpler
7. Choose black boxes carefully
8. Invent things that you would want to use yourself
9. Balance user suggestions with observation and participatory processes
10. Iterate, iterate—then iterate again
11. Design for designers
12. Evaluate your tools

CSTs vary in their application of the previously discussed objectives, roles, and principles. Of the many examples in the literature, I highlight the skWiki system, which was designed to support collaborative sketching (Zhao et al., 2014). Its main form of support borrows from the mechanics of code versioning systems, creating a revision history of sketches. This means that users can create sketches of their own (depicting, for example, a system's UI), as well as see sketches done by other users. Each set of changes they make on their sketch advances its history. Similarly to code versioning approaches, a user can only modify someone else's sketch by creating branches, which can be later merged. Users can also rollback to previous versions and start working from there. This is in line with the principle of supporting exploration (Resnick et al., 2005), as users can try out different configurations without fear of ever losing any progress they make. The versioning feature therefore acts as a ski, allowing a new form of interaction that would otherwise not be possible, and as a pen pal, supporting creative collaborations.

Creativity Research Summary

In this chapter section, I reviewed current literature on creativity with special emphasis on idea generation and its associated models—SIAM and Brown's Matrix. These models describe idea generation as being enabled by each ideator's unique cognitive model, that is, their knowledge and associations they make. I also discussed the relevance of idea combination and development, as well metrics for evaluation creativity. I concluded with a discussion of CSTs and their underlying design principles and approaches for supporting creativity. Much of this research is concerned with an individual or small group level. I now survey the crowdsourcing literature with a particular focus in how crowds have been used in the context of large-scale creativity.

Large-Scale Creativity

Crowdsourcing and Microtasks

With the possibilities brought about by the internet, both research and industry started exploring increasingly larger group sizes for many different types of tasks—an approach known as crowdsourcing (Howe, 2006). While definitions can vary, we can say that crowdsourcing happens when a *requester* (an individual or organization) proposes the undertaking of a task to a *crowd* (a large group of people) (Estelles-Arolas & Gonzalez-Ladron-de-Guevara, 2012). For example, a requester in need of audio transcription services could propose this work to a crowd such as the one available in Amazon's Mechanical Turk. There can be variations, however, as the initiative could also rise from the crowd itself, or the results could be crowdsourced passively rather than actively (Bigham, Bernstein, & Adar, 2015).

Since its conception, crowdsourcing has seen some notable results. One of the most notable examples, Wikipedia[4], tapped into the intelligence of the crowd to create an encyclopedia with over 5.3 million articles in English, and which features reasonable error rates when compared to leading physical encyclopedias (Giles, 2005). And while there have been cases of notable errors, intentional or otherwise, the fact that they can be instantly corrected also weighs in Wikipedia's favor. Other notable accomplishments can be found in the domains of citizen science (Kanefsky, Barlow, & Gulick, 2001) and the creation of complex content such as animations and product design (Retelny et al., 2014; Valentine et al., 2017). Other useful results stemming from crowds are found in popular

[4] http://www.wikipedia.org

apps such as Waze[5] (crowdsourcing traffic information) and Duolingo[6] (crowdsourcing translation).

One common approach in crowdsourcing, and the most relevant for this work, stems from the affordances of Amazon's Mechanical Turk (MTurk) platform. This approach involves breaking down tasks into the small units—*microtasks*—and assigning them to crowd members. These microtasks are connected by intricate *workflows,* which at the end generate results that attempt to be at least comparable to what a single expert could produce. Perhaps the best example for this approach exists in the CrowdForge framework (Kittur, Smus, Khamkar, & Kraut, 2011). This framework defines three main types of tasks: *partition*, which breaks down larger tasks into smaller ones; *map,* in which the tasks are performed; and *reduce,* in which the results from the map tasks are joined together. They exemplify this through a workflow for writing Wikipedia articles. Their partition task asks workers for an article outline. The map tasks had multiple workers submitting facts for one item of the outline in the partition task. Finally, the reduce task asked workers to write paragraphs for each outline item based on all the facts collected on the map tasks. They found this workflow yielding better results than when workers were asked to write the entire article by themselves, and comparable results to an existing simplified Wikipedia article on the same subject. These results translate to different domains, as similar microtask workflows have been developed and successfully deployed for tasks such as science journalism (Kittur et al., 2011), taxonomy creation (Chilton,

[5] https://www.waze.com
[6] https://www.duolingo.com

Little, Edge, Weld, & Landay, 2013), design feedback (Luther et al., 2015), collaborative writing (Teevan, Iqbal, & von Veh, 2016a) and many others.

The common thread within this body of work is that microtasks functioned as a form of scaffolding, allowing workers with limited attention or expertise to contribute towards goals that they might not be able to achieve, at least with acceptable quality, by themselves. These successes stem from some of the affordances and benefits of microtasks, which, compared to larger tasks, have been found to yield less cognitive load, fewer errors, and to be more robust to interruptions (Cheng, Teevan, Iqbal, & Bernstein, 2015), can promote learning (S. Dow, Kulkarni, Klemmer, & Hartmann, 2012; Luther et al., 2015), and can be easily integrated into computer code (Little, Chilton, Goldman, & Miller, 2010). However, microtasks may also take longer to complete (Cheng et al., 2015; Chilton et al., 2013), suffer from issues of lack of context and blocking (Teevan et al., 2016a), or simply be utterly boring (Siangliulue et al., 2016; Teevan et al., 2016a). Crowd markets such as Mechanical Turk also elicit ethical discussions such as fair compensation for workers (Bigham et al., 2015). Therefore, the choice of using microtasks (and the crowd markets that support them) as the method for crowdsourcing a task must be weighed against these factors, with microtasks being seemingly more useful to users who would otherwise not be able to complete the entire tasks due to lack of knowledge, motivation, or time.

Crowdsourced Creativity

While the work described above largely aims at eliciting the intelligence of the crowd, other efforts started to focus on their creativity, hoping to fulfill the creative potential of distributed and diverse groups (Fischer, 2005). In this space, there are notable names such as InnoCentive, OpenIDEO, and Threadless . These initiatives aim at eliciting creative products—from t-shirt designs to the development of innovative new materials—through an open call initiated by an interested party. Creativity can also be organically crowdsourced. For example, one could consider initiatives such as Open Source Software to be a form of crowdsourced creativity, as users engage in creative activities (e.g. developing new software features) around a common interest.

Much of crowdsourced creativity and CST research focuses on the ideation phase, mostly through crowd-powered brainstorming sessions. Support during ideation usually employs a peripheral task facilitation model, where the ideas generated by the crowd (or properties of the problem) are sent to a third-party individual or group, who processes those ideas in some way. The results are then sent back to ideators, with the goal of supporting interventions that can increase idea generation performance. For example, Yu, Kittur, & Kraut (2014a) had workers generate problem schemas—generalizations of solutions to a problem—and subsequently showed these schemas to ideators in order to enhance their idea generation performance in a subsequent study. The results from these tasks can also be used by researchers to measure the effects of their interventions (e.g. using it to determine the intervention effect on the originality or usefulness of ideas).

These peripheral microtasks can be organized within four broad categories: 1) Rating, where users are ask to rate ideas in terms of their novelty, value, or similarity to other ideas; 2) Combination, which presents users with a few ideas and asks them to combine their features into a new idea; 3) Inspiration design, in which users are given instructions to generate some kind of artifact that can be used as an inspiration for ideators; and 4) Problem abstraction, which asks users to generate some features related to the description of the problem. Table 1 enumerates the tasks within each category. These tasks will output information that can be used as the basis for creativity-boosting interventions. For example, rating tasks can classify ideas as high/low creativity, or group them into clusters of similar ideas. With this information, a system can then present users with diverse sets of ideas, which can improve the diversity of idea generation (Siangliulue, Arnold, et al., 2015).

Table 1

Peripheral task types and examples

Task type	Domain
Rating	**Rate an idea in terms of novelty/originality and quality/practicality/value** (Chan et al., 2016a, 2016a; Siangliulue, Chan, et al., 2015; Yu & Nickerson, 2011)
	Choose the most similar idea to a seed (Siangliulue, Arnold, et al., 2015)
	Rate similarity of ideas (Siangliulue, Arnold, et al., 2015)
Combination	**Combine aspects of two different designs** (Yu & Nickerson, 2011)
Inspiration design	**Generate inspirations (thought provoking questions, insights, theme)** (Chan et al., 2016b)
	Find inspiring images or descriptions (Yu, Kraut, et al., 2016)
	Find analogous ideas (Yu et al., 2014b)
	Find ideas within a domain (Yu, Kittur, & Kraut, 2016)
Problem abstraction	**Extract features from problem description** (Yu et al., 2014b)
	Ask workers to generate/summarize constraints (Yu, Kraut, et al., 2016)
	Suggest experts that could help with the problem (Yu, Kittur, et al., 2016)

Another way of looking into current crowdsourced brainstorming research is by organizing it around the method of inspiration used to enhance ideation. At the most basic level, this has been done by simply showing ideators other ideas. The approaches differ in how they choose the ideas or how they show them. For example, selecting a set of diverse or creative ideas can improve their effects over random sets of ideas (Siangliulue, Arnold, et al., 2015). The timing and delivery method of these inspirations can also affect their efficacy, showing benefits to giving users a choice of when to receive inspirations or in employing a smart strategy for choosing the right moment to do so (Siangliulue, Chan, et al., 2015). Finally, increasing the attention to the inspirations ideas, such as by asking questions about the inspirations, can also improve performance under certain circumstances (as described in Chapter 4, as well as Girotto et al., 2017).

Since simple exposure to other ideas can bring its own set of issues (e.g. fixation (Jansson & Smith, 1991)), others have proposed ways of inspiring users through abstractions of other ideas or features of the problem. For example, previous work has found some advantage to using machine-generated abstractions of others' ideas as an inspiration (Chan et al., 2016a). Alternatively, crowd workers can be used to identify and generate schemas to be used as inspirations (Yu et al., 2014a, 2014b). Another way in which abstractions can be used is through real-time facilitators. This was tested by Chan, Dang, & Dow (Chan et al., 2016b) through their IdeaGens system. They found that by using stimulating strategies such as simulations (asking ideators to imagine scenarios), facilitators can improve ideator's fluency and creativity. Features of the ideation problem can also contribute, such as by identifying domains of expertise relevant to it, or presenting ideators with constraints (Yu, Kittur, et al., 2016; Yu, Kraut, et al., 2016).

A final approach considered here is that of highly structured human-powered processes. For example, it has been shown that a human-powered genetic algorithm, in which ideas are mixed and selected through several iterations, can result in greater creativity of later ideas (Yu & Nickerson, 2011). Perhaps even more structured, the BlueSky system employs a crowd-powered algorithm to more evenly contribute to the solution space and reduce duplicates (Huang & Quinn, 2017).

Considering the research reviewed above, a few limitations with the current support approaches are made evident. The first one is that support is generally passive. Research has shown that attention is a key factor for external stimulus to affect ideation (Dugosh et al., 2000). Passive forms of support, therefore, may not be eliciting enough attention to the inspirations, and may therefore be curbing their effects on ideation. The second is that the support is the same for every ideator. The SIAM model presents idea generation happening around LTM structures that are unique to each ideator (Nijstad & Stroebe, 2006), and the matrix model even more explicitly describes individual ideators as unique transition probability matrices (Brown et al., 1998). This means that the effect support has on ideators depends on *which ideas* the support exposes ideators to. Therefore, current support approaches may not be efficient due to lack of attention and may not be appropriate because they do not consider users' unique cognitive structures. As described in later chapters, this dissertation aims to address those issues through active inspirations (Chapter 4) and an adaptive system (Chapter 5).

Crowd Creativity Assessment

An essential but challenging factor in crowd creativity research is the evaluation of ideas. Metrics reviewed previously, such as fluency, are relatively straightforward to scale up to hundreds of ideators, but others prove to be difficult. For example, the consensual assessment method (Hennessey et al., 2011) requires independent judges to rate all ideas relative to each other—sometimes in more than one dimension. This is likely to be infeasible in a large-scale context. To counter that, researchers have turned either to crowdsourced evaluation or to automatic methods of comparison.

Crowdsourced methods mainly use Mechanical Turk workers to evaluate the originality and usefulness of ideas. These can be seen primarily in the *rating* tasks described in Table 1. The process is fairly similar across the different studies: a worker is shown a number of ideas (between 12 and 20) and is asked to evaluate those ideas in terms of their novelty/originality, as well as their quality/practicality/value (Chan et al., 2016a, 2016a; Siangliulue, Chan, et al., 2015; Yu & Nickerson, 2011). Although this is similar to the process of consensual assessment (Hennessey et al., 2011), judges (usually from Mechanical Turk) may either lack expertise to judge or may simply yield an assessment that is too imprecise to be usable. These points could affect the final reliability of the judgements, which is a core requirement for its validity (Hennessey et al., 2011). In my experience, following a similar methodology to the one outlined above, I found very low agreement between raters. Similar issues were also reported by Chan et al. (2016a). Therefore, extracting subjective metrics, such as these from crowds can be extremely challenging, and the results elicited from crowd markets must be taken with care.

Conclusion

In summary, the research reviewed in this chapter highlights the nuances of creativity, represented by the many different perspectives, levels, and approaches to examining this important phenomenon. From the richness of this field, I focused some more on the describing the process of idea generation, commonly known through the brainstorming technique. This includes a description of the cognitive processes underlying idea generation, as well as the importance of combination and iteration for the creativity of the output. I also described the different ways in which creativity can be assessed, as well as how tools can be built to support it. From there, I described how research (e.g. through microtask markets such as Mechanical Turk) and practice (e.g. through initiatives such as open source or innovation challenges) have attempted to tap into a larger-scale creativity, involving hundreds or thousands of ideators in coming up with ideas towards a common goal.

The next chapter extends this literature by diving deeper into the current practices of online communities. Particularly, I look into Free and Open Source Software communities. They engage in creative activities such as coming up with new features for their projects. But since these communities are distributed, they must use certain online collaboration tools to engage. Under the guidance of the literature on CSTs describe above, I analyze these collaboration tools and their appropriateness for supporting these creative collaborations.

CHAPTER 3

BOTTLENECKS IN CURRENT COMMUNICATION CHANNELS

Online communication technologies provide flexibility to diverse distributed teams and communities, enabling collaboration across cultures and geographical locations. Such a context can enable creativity to flourish (Dennis & Williams, 2003; Fischer, 2005). When these groups develop creative ideas, that is, those that are novel and useful (T. Amabile, 1988; Hennessey & Amabile, 2010), there is great potential for innovation—the successful implementation of those ideas (T. Amabile, 1988). Nonetheless, many distributed groups engage in creative collaborations (e.g. brainstorming sessions) using channels such as discussion boards or mailing lists, which were not designed for that purpose. And since the affordances of communication tools influence how the communication happens (Y.-C. Wang, Joshi, & Rosé, 2008), these channels may shape discussions in ways that are not conducive to creativity, ultimately affecting the level of innovation these distributed groups produce.

Therefore, in this chapter I explore the following question: **in which ways are the affordances of the communication channels used by online distributed communities hindering their creativity?** This analysis contributes to my overarching goal of supporting large-scale creativity by identifying breakdowns in the collaboration of online distributed communities. As a framework for analyzing this question, I rely on existing research into Creativity Support Tools (CST). CSTs are tools that aim to support creativity by enabling greater exploration, experimentation, and combination of ideas (Shneiderman et al., 2006; See Chapter 2 for more information). Principles for designing

CSTs have emerged throughout the research literature. In this chapter, I survey these principles and apply them as guiding heuristics for an analysis of the channels in Free and Open Source Software communities (FOSS), using three methods: a heuristic analysis of the channels' UIs, a content analysis of discussions in those channels, and interviews with members of distributed groups.

I chose to focus on FOSS for the following reasons: 1) FOSS communities constantly engage in public creative activities, such as discussing the design of a new feature or where the project should be heading. 2) FOSS members are likely to have the knowledge and motivation needed for creativity to occur—two requirements for creativity (T. Amabile, 1988); 3) FOSS projects are diverse in domain, tools used, and duration of projects; and 4) FOSS projects are important to society and enterprises. For example, the most used web server is Apache, an open source project (Netcraft, 2018). Nonetheless, even though I focus on FOSS, I also expect these contributions to expand to other distributed groups, as the channels I evaluate are not exclusive to FOSS.

In this chapter, I make the following contributions:

- A set of heuristics from the CST literature that can be used to guide the analysis of channels for creative collaboration;
- A description of common creativity breakdowns found in public FOSS discussions;
- A description of design elements in common communication channels that may bottleneck the creativity in FOSS collaboration and promote these breakdowns;

These contributions have implications for the design of communication channels and CSTs for FOSS communities as well as for other distributed groups that employ similar channels for collaboration.

The remainder of this chapter is as follows. I start with a description of FOSS projects and the channels they use, as well as a listing of the communities we evaluated. I then review research on CSTs with the purpose of consolidating a set of design heuristics to guide our analysis. Then, I define the methodology used for the three types of analyses employed in this chapter. In sequence, I describe the results in terms of the issues identified through the three methods. I then conclude with a discussion of the results, their implications, and the limitations of our approach.

FOSS Communication Channels

FOSS projects are defined, among other characteristics, by the openness of their source code which anyone can download and modify (Open Source Initiative, 2007), thus enabling anyone to contribute back to the project. This development model has been compared to a bazaar, an environment of apparent chaos and frequented by diverse crowds, but from which stable systems emerge (Raymond, 1999). This diversity and openness offers great potential for creative collaborations (Fischer, 2005). In fact, open source contributors show a high level of motivation and expertise, and often experience a high sense of creativity and flow while contributing to open source communities (Lakhani & Wolf, 2003). This makes FOSS communities a fertile ground for creativity (T. Amabile, 1983), and consequently innovation. In fact, many enterprises run on open source software such as Linux and Apache. Nonetheless, their collaboration processes need to be conducive to creativity if that potential is to be fulfilled (T. Amabile, 1983, 2012). A classic example is brainstorming, which aims at helping people generate more ideas by adhering to simple rules such as withholding criticism (Osborn, 1963). However, in distributed communities such as those in FOSS projects, these creative collaborations must happen through online communication channels such as discussion boards, mailing lists, or bug trackers. The affordances of these environments can negatively shape how these creative processes happen. I now briefly discuss these communication channels.

Many FOSS projects communicate through *mailing lists*, the oldest of the channels surveyed here. Mailing lists employ a push pattern of communication, which may yield a stronger perception that the messages sent through it will be seen (Zhang, Ackerman, & Karger, 2015). However, this same feature of mailing lists may hinder some from using them out of a fear of spamming the group or saying something “stupid” to a large audience (Zhang et al., 2015). Previous research also highlights issues with searchability of the mailing list archives, as well as difficulties in performing work over a shared document (Saeed, Rohde, & Wulf, 2011). While their simplicity can also be seen as an advantage, other channels may provide more appropriate affordances or incentives than those of mailing lists, causing some migration away from them (Vasilescu, Serebrenik, Devanbu, & Filkov, 2014).

Alternatively, groups can also collaborate through *discussion boards*. Discussions boards often differ in their implementation, which influences how the discussions unfold. For example, some afford parallel discussions, through mechanisms such as replying to specific messages, while others follow a strict sequential progression. In the former kind we find that responses are less constrained by a chronological factor (Y.-C. Wang et al., 2008), but there can still be issues of a lack of responses or interaction with responses, as well as a lack of meaningful interactions between posters (M. J. W. Thomas, 2002). Furthermore, a newcomer who arrives at a long discussion may be less likely to participate (McInnis, Murnane, Epstein, Cosley, & Leshed, 2016).

Finally, in addition to domain agnostic channels such as mailing lists and discussion boards, communities may also employ collaboration tools that are more specific to their domain. For example, Free and Open Source Software (FOSS) communities heavily collaborate using *Code Versioning Tools (CVS) or bug trackers,* of which GitHub is perhaps the most notable example. GitHub has been described as a catalyst of new contributions by making it easier for newcomers to contribute to projects (McDonald & Goggins, 2013; Vasilescu, Filkov, & Serebrenik, 2015). The fact that collaboration revolves around coding allows users to infer skills and needs of other developers due to their history of actions (Dabbish, Stuart, Tsay, & Herbsleb, 2012). Nonetheless, the code-centric affordances of GitHub and other CVS tools may push away non-technical users (Zagalsky, Feliciano, Storey, Zhao, & Wang, 2015). And although GitHub has social affordances, there is evidence that much of the social interactions happens outside of it (Wu, Kropczynski, Shih, & Carroll, 2014).

While the brief review above outlines some general implications of the design of those communication channels, I hope to further our understanding of those channels by examining them through the lens of creativity support, detailing how specific design choices can affect creativity. Therefore, I analyzed a set of twelve FOSS projects that rely on those communication channels. The projects were chosen to cover a broad range of domains to also capture participation from non-developer contributors. I also surveyed communities with different management strategies, from those managed by the community and its leaders to those that were managed by companies or organization. Finally, I sought to cover both long-active projects and relatively new projects, as their choice of communication channel depends on the popular tools available at the time the project is created. A summary of the projects can be seen on Table 1.

For each project, I looked for the most active channel used by members to discuss new features. The communities here examined used mailing lists, discussion forums, and CVS/bug trackers to collaborate on designing new features.

Table 1

List of FOSS projects that were analyzed

Project	Domain	Management	Channel Type	Average Interactions per Thread
Audacity	Audio	Community	Forum	41.6
Brackets[7]	Software	Company	Mailing List	11
Drupal	Software	Association	Bug Tracker	28.5
FlightGear	Gaming	Community	Forum	14.9
FreeCAD	CAD	Community	Forum	35.9
GIMP	Image	Community	Bug Tracker	24.7
LibreOffice	Office	Foundation	Bug Tracker	22
Minetest	Game	Community	Bug Tracker	16.5
Moodle	Education	Company	Forum	27.1
MuseScore	Music	Community	Forum	17.2
VS Code	Software	Company	Bug Tracker	16.3
Zotero	Science	Community	Forum	38.3

[7] When the discussions were collected in September 2015, I found the Brackets mailing list to be the most active channel for discussing new features. By the time this chapter was written, the channel's popularity seemingly died down. Nonetheless, I kept it in our analysis as it is an example of a different type of communication channel.

CST Design Heuristics

There are several sources for design principles for CSTs, some of which I have discussed on Chapter 2 of this dissertation. Perhaps most significantly, a 2005 NSF-Sponsored workshop yielded a list of 12 design principles for tools that support creative thinking (Resnick et al., 2005; Shneiderman et al., 2006). These include items such as supporting exploration and different styles of creative thinking. However, there are other lists of important principles to be followed, either through comprehensive lists like the one produced by the workshop (e.g. Herrmann, 2009; Selker, 2005) as well as more focused ones, especially as they relate to specific implementations (e.g. Zhao et al., 2014). Nonetheless, there is much overlap between the various sources of CST design principles, and some principles cannot be directly applied to the evaluation of an interface (e.g. "*Invent things that you would want to use yourself*", Shneiderman et al., 2006). Given that one of the goals for this chapter is to perform a heuristic evaluation of communication channels in terms of their creativity support affordances, I consolidated the design principles from these various sources into a list of six heuristics. I note that many of the lists mention principles related to usability of systems. While I acknowledge the importance of usability for creativity support, I omit these principles to focus on those more directly relevant to creativity support.

1. Support for divergent thinking (DT)

CSTs should support divergent thinking, which is the generation of many diverse ideas or alternative solutions. This has been expressed in different ways, such as the need for CSTs to support divergent thinking (Farooq, Carroll, & Ganoe, 2008; Zhao et al., 2014), exploration of alternatives (Arias, Eden, Fischer, Gorman, & Scharff, 2000; Hewett, 2005) and different paths (Resnick et al., 2005), brainstorming (Selker, 2005), and variation (Herrmann, 2009). This can be supported, for example, through preservation of minority ideas and dissent (Farooq et al., 2008; Farooq, Carroll, & Ganoe, 2007), a library of macros and analogs (Hewett, 2005), and allowing simultaneous representations for comparison (Hewett, 2005). In fact, supporting divergent thinking is often the main focus of CSTs (Gabriel et al., 2016; K. Wang & Nickerson, 2017). In the creativity literature, divergent thinking is commonly examined through famous techniques such as brainstorming, in which the goal is to generate as many ideas as possible (Osborn, 1963), or psychometric measures such as the unusual uses test (Guilford et al., 1958), in which individuals are asked to come up with as many unusual uses for an object as possible. A high level of divergence is a predictor of the creative potential of individuals or groups (Runco, 2008).

CSTs can improve divergent thinking by aiding ideators in coming up with more ideas towards a given problem. For example, tools such as CreaCogs-OROC can automatically generative alternative uses to objects (Olteţeanu & Falomir, 2016). These different uses could inspire or extend the divergent thinking of a human. Another such environment is found in the IdeaExpander tool, which augments brainstorming by processing the conversation and displaying stimulating images (H.-C. Wang, Cosley, & Fussell, 2010). Alternatively, simpler approaches such as supporting the presence of a facilitator during ideation can already provide significant support for divergent thinking (Chan et al., 2016b).

2. Support for convergent thinking (CT)

This is the often-neglected counterpart to divergent thinking, despite its importance. In the literature, this has been expressed as the need for CSTs to support convergent thinking or convergence (Farooq et al., 2008; Herrmann, 2009), a critical evaluation of perspectives (Farooq et al., 2007), reflexivity (Farooq et al., 2008, 2007), recording of rationale (Arias et al., 2000), and simulations and analysis (Arias et al., 2000). While divergent thinking aims at expanding the solution space by generating multiple alternatives, convergent thinking focuses on reducing this solution space through reflection and rationale. Some examples of convergent activities are grouping items, seeking accuracy, and seeking the best answer (Cropley, 2006). In fact, some operationalize creativity as something both new and useful (Hennessey & Amabile, 2010), the latter of those being directly related to convergence. Divergence without convergence can lead to unrestrained, disastrous ideas (Cropley, 2006).

Support for convergent thinking is exemplified in the work of Davis and colleagues, who explored the use of a system to improve hobbyists cinematographic performance by presenting them with a simulated analysis tool that could warn them of any cinematographic rules they broke (Davis et al., 2013). By pointing out rule violations, this tool is virtually reducing the solution space for their video productions, and thus can act as a form of convergent thinking support. In a similar manner, the KID Design Environment supports users in designing a kitchen layout through a catalogue of rules, which can accuse problematic situations in the design (Nakakoji, 2006).

3. Support for shared material (SM)

CSTs should support the sharing and manipulation of different types of materials rather than only text. This has been expressed as the need for CSTs to support malleability of shared material (Herrmann, 2009), integration of communication with shared materials (Herrmann, 2009), easily transferable rich media (Zhao et al., 2014), and low-cost modifiable models (Arias et al., 2000). These materials should be easily accessible through multiple paths, including search mechanisms (Hewett, 2005). The purpose of supporting non-textual materials, such as sketches or prototypes, is that they can communicate some ideas better than text could. Furthermore, some materials such as 3D models can enable others to simulate the ideas, further aiding in convergent thinking.

This principle is clearly exemplified in the skWiki system (Zhao et al., 2014). skWiki allows users to produce low cost sketches that can be viewed and modified by other users. Alternatively, while not strictly a CST, versioning tools such as GitHub, which is discussed further in this chapter, also adhere well to this principle by allowing discussions to be grounded on the shared material (e.g. a given line of code within a commit), which can be changed through further commits.

4. Support for shared understanding (SU)

CSTs should support the development of a shared understanding of the problems and the proposed solutions. This has been expressed in the literature as the need for supporting the larger picture (Herrmann, 2009), the development of shared objectives (Farooq et al., 2008), and the development of shared understanding through artifacts (Arias et al., 2000). Shared understanding has been described as key for creative collaborations, especially as the challenges become increasingly interdisciplinary and individuals need to collaborate in order to tackle them (Arias et al., 2000). Without shared understanding and goals, teams may fail to accomplish their plans (Bittner & Leimeister, 2014; Hinds & Weisband, 2003).

The BRIDGE system, described by Farooq and colleagues (Farooq et al., 2007), provides a good example of how this principle can be supported. This system supports a shared understanding of group participation in two ways: 1) it can provide users with summaries of discussions, which can help a new or returning member to be aware of the discussion that took place while he or she was absent; and 2) it supports structured updates, that is, small snippets of text describing *who* is working on *what*, allowing community members to avoid duplicated work and helping them to better coordinate their efforts.

5. Support for collaborative and iterative processes (CI)

CSTs should support the creative process through iterations. In the literature, this has been described as the need for CSTs to support iteration (Resnick et al., 2005), planning (Farooq et al., 2008), a revision history (Zhao et al., 2014), multiple configurations of the work environment (Hewett, 2005), the logging of processes and intermediate results (Hewett, 2005), and a history of changes (Selker, 2005). But these iterative processes are likely to be collaborative, and thus tools should also support collaboration (Herrmann, 2009; Resnick et al., 2005; Selker, 2005; Zhao et al., 2014). The usefulness of collaboration for creativity is quite well-known, although the way people interact together needs to be carefully designed, as several processes such as production blocking and fixation may lead groups to perform worse than individuals working in isolation (Diehl & Stroebe, 1991). As for iterations, it is a key component to develop the maturity of ideas. For example, it has been found that combinations of ideas may only yield creative ideas after they have undergone some iterations (Chan & Schunn, 2015).

Perhaps this principle is most visible in the skWiki tool (Zhao et al., 2014), already described above. One of its main features is the path model, which describes each sketch as a set of operations over a previous sketch, akin to code versioning tools. This allows each sketch's history to be iterated over without fear of losing the original work. This model also allows for operations such as branching or merging different versions. This is again parallel to GitHub, in which users can work in parallel within their individual branches, and merge them at a later point in time, with a complete history of changes behind the latest version. This kind of support allows users to collaborate in a highly iterative environment, in which artifacts can be changed without fear of lost information.

6. Support for group diversity (GD)

CSTs should support a diverse set of users. As described in the previous principle, CSTs are likely to be used in a collaborative context, and therefore they should support collaboration. But this collaboration is also likely to happen among peers with diverse background. Therefore, people need to be supported in their individual traits. In the literature, this has been described as the need for CSTs to support many styles (Resnick et al., 2005), as well as leverage cognitive conflict and minority dissent (Farooq et al., 2008). Perhaps more emphatically, the NSF workshop report describes the need for tools to support a “low threshold, high ceilings, and wide walls (Resnick et al., 2005)”. By this, they mean that CSTs should support novices (low threshold), experts (high ceilings), and a wide range of explorations (wide walls). Diversity may bring about different skills and knowledge, which can improve performance of the group (Fischer, 2005; McLeod, Lobel, & Cox Jr., 1996; Milliken, Bartel, & Kurtzberg, 2003). However, the group interaction has to be carefully thought out, as the diversity may also ask as a source of contention or miscommunication (Milliken et al., 2003).

The work of Davis and colleagues (Davis et al., 2013) described above is a good example of this form of support. It encourages novices to explore the space, supporting them with knowledge on cinematographic rules they might be advised to follow. It therefore has a low threshold. Nonetheless, the ceiling remains high, as experts would not be limited by this form of support.

Method

To explore the research question of how FOSS communication channels may be hindering their creativity, I carried out three different types of analysis: a heuristic analysis of the channels’ interfaces, a content analysis of discussions, and interview with FOSS contributors. These analyses were done in conjunction with two other graduate students (doing research in the area of HCI and CSCW). I now describe each method in more detail.

Heuristic Analysis

The heuristic analysis was carried out by three analysts (following the baseline recommendation of three evaluators, Nielsen & Molich, 1990). To conduct the heuristic analysis, we compiled a list of the main communication channels and their URLs for each of the projects in Table 1 of this chapter. If two projects shared the same channel (e.g. GitHub), the channel was included only once in the list. We then met to discuss the heuristics, their rationale, and application. The goal for these meetings was to ensure a shared understanding of the heuristics between the three of us, including how issues with their compliance could look like.

We independently analyzed each of the channels, compiling a list of issues based on the heuristics. We occasionally met early on the process to ensure correct understanding of the heuristics. Each one of us compiled a list of issues, each issue including: 1) a short description of the issue; 2) the relevant heuristic; 3) the channels/projects in which the issue occurred; 4) a longer description of the issue, including screenshots if applicable. After all evaluators concluded their analysis, their lists were sent to me, who compiled a shared list of issues. This yielded a list of 16 issues, with 3 being identified by all evaluators, 3 being identified by two of them, and 10 being identified by only one.

Content Analysis

The second analysis focused on the content of discussions that happened in the channels, aiming to see how the issues identified in the heuristic analysis may manifest themselves in the discussions. To perform this analysis, I collected 10 discussion threads from the 12 projects listed in Table 1. I focused on collecting discussions around new feature suggestions, or changes to some aspect of the project. I chose these kinds of discussions as they more closely resemble creative activities such as idea generation sessions and afford the generation of multiple alternatives before coming to a decision. Table 1 shows the average number of messages in each thread for each community.

Discussions were analyzed through a general inductive approach (D. R. Thomas, 2006), using the CST design heuristics as the overall evaluation goals, while letting subcategories emerge from the data. Initially, I went through a subset of the discussions (~10%) and developed subcategories. A second pass on this subset was performed to merge or exclude these categories into a smaller and more cohesive set. To increase the trustworthiness of this coding, another member of the research team independently coded the same subset, occasionally meeting with me to discuss and merge the results. These subcategories were then used by me to code the rest of the data.

Table 2

Distribution of content coding across communities and heuristics

Project	DT	CT	SM	SU	CI	GD	TOTAL
Audacity	4	6	6	27	0	8	51
Brackets	10	5	2	9	0	0	26
Drupal	6	2	0	11	1	0	20
FlightGear	3	1	2	26	1	3	36
FreeCAD	9	6	10	22	0	5	52
GIMP	3	1	9	38	2	6	59
LibreOffice	7	7	1	37	1	3	50
Minetest	5	3	1	3	0	3	15
Moodle	16	4	8	55	4	8	95
MuseScore	2	1	8	35	0	10	56
VS Code	3	2	3	7	0	0	15
Zotero	3	7	3	61	0	6	80
TOTAL	65	45	53	331	9	52	555

In total, I coded 555 situations throughout the 12 communities analyzed. Table 2 summarizes their distribution across communities and heuristics.

Contributor Interviews

Finally, I conducted interviews with contributors of the projects listed in Table 1. The purpose for these interviews was to provide further insight into the issues uncovered in the content analysis. To recruit participants, I posted announcements in the projects' discussion pages. I explained the purpose of the project and asked those interested in participating to fill out a small contact form, used for screening potential participants, as well as for scheduling purposes. Interviews happened online over the course of a week, and with each lasting for about one hour. Interviews were semi-structured and focused on their experiences when discussing new features. Participants were compensated with a $10 Amazon gift card. In total, I interviewed four contributors from three different projects (P1 and P3: Zotero; P2: FreeCAD; P4: MuseScore), with different levels of experience and backgrounds. Their years of contribution to open source ranged from 5 to 10 years. Analysis of interviews was informal, with one research member listening to the audio and extracting quotes that related back to the six heuristics.

Results

I related the design elements (uncovered in the heuristic analysis) to the issues (uncovered primarily in the content analysis) by looking at the content coding and reasoning over which UI design elements could be contributing to those issues. Some design elements (e.g. no way to browse all the materials shared in a discussion) ended with no corresponding issue, just as some issues (e.g. animosity in conversation) had no corresponding UI affordance. Therefore, I excluded those. The remaining issues are presented in Table 3.

Table 3

Relationships between Design Principle, Design Element, and Issues

Design Principle	Design Element	Issues
Divergent thinking (DT)	1) Initial topic sets the tone for the discussion	Discussions start and center around well-defined ideas
	2) Ideas coexist with chronological discussions	Ideas are lost in discussion
Convergent thinking (CT)	1) No support for parallel discussions	Unproductive discussions
	2) No way of indicating support	Support is expressed in non-intuitive ways
		Convergence starts early
	3) No mechanism for concluding discussion with actionable outcomes	Discussions don't arrive at a clear conclusion
Shared material (SM)	1) You cannot share many types of files	Users resort to laborious representations
	2) Shared materials are not editable, interactive, and are disconnected from discussion	Difficulty in communicating ideas
Shared understanding (SU)	1) Related discussions are not tracked together	Redundant discussions
		Lack of knowledge of related discussions
	2) No summarization of conversations	Users will not read previous messages
Collaborative, Iterative Process (CI)	1) No versioning for shared materials	Users must manually notify and update others
	2) No integration with synchronous collaboration	Users must manually relay information back to asynchronous channels
Group diversity (GD)	Technical language and paradigms	Non-technical users may be confused or discouraged to participate

For each heuristic in Table 3, I also list the design elements and their associated issues. Throughout this discussion, I refer to specific design elements by their heuristic code and design element number. For example, CT-2 refers to the second design element in the convergent thinking heuristic (*No way of indicating support*).

Bottlenecks in Divergent Thinking

The first issue revolves around how the discussions start. Most discussions start around a well-defined idea (**DT-1**). Naturally, the cause for this is not necessarily on the channel itself: most ideas come from personal needs that emerged while using the software. However, this focuses the discussion around that particular idea rather than on addressing the underlying issue. This can limit divergence in two ways. First, it may focus the discussion on the specifics of that idea rather than on generating alternatives, or even for others to simply manifest their support for that idea. And second, it may fixate others into the features of that idea (Smith, 2003). For example, in one discussion, a user expresses his curiosity at the similarity of all the ideas: "*it is interesting how much we were thinking the same when we started to explore options and took the challenge*".

The second issue with divergence is that ideas and discussions about those ideas share the same space (**DT-2**). There is no hierarchical or visual differentiation between them. In the content of discussions, we identified that frequently ideas got lost in the discussion. In fact, P2 expressed concern with this: "*It's just idea, after idea, after idea, and conversation. They're just piled on after the other. How can we use this?*" Throughout many of the discussions, we've identified instances in which ideas were mentioned in the middle of a discussion but were never acknowledged. It is possible that people may have read it and simply chose not to comment. However, as the discussion continues, these ideas become lost in its midst. But problems can happen when they are posted towards the end of the discussion as well. Many users arrive late into the discussion. When they add their ideas, they express a feeling of being late into the process (e.g. "*I'm joining the conversation late*", Zotero), even apologizing for it: "*Hey sorry for jumping into this thread belatedly*" (Brackets). The channels' affordances should avoid this perception, as users who could not participate earlier, for various reason, may hesitate to share their diverging contribution so late in the process.

Bottlenecks in convergent thinking

The first issue we identified for convergent thinking is that of no support for parallel discussions (**CT-1**). This relates somewhat to the divergent thinking issue of ideas coexisting with discussions, but the focus here is on the discussions themselves. Posts are organized chronologically, with posts visualized one after the other. The implication is that even if different ideas have been suggested, it is difficult to distinguish between parallel discussions. In fact, we often found that discussions tend to be confusing and unproductive. For example, after a user expressed support for an idea, someone replied with a query about which part of the discussion he was supporting: "*Is that +1 for having a "sone scaled view" or +1 for a sone scale view being the default view*" (Audacity). In LibreOffice, another user says that the "*goal of this issue has become muddled*" after another user mentions how he is confused about which of the several proposed options is being asked.

The discussion channels often lacked mechanisms for succinctly expressing agreement (**CT-2**). This resulted in many users seeking to express support however they could. Generally, they would simply write a variation of "*I like that idea*" (Brackets). Others suggested adding themselves to a list of users to be notified whenever changes happened to the thread, saying that "*This would give an idea of bugs popularity. Less accurate than votes, but an idea*" (LibreOffice). Therefore, this impedes their ability to quickly estimate popularity of ideas. The other issue that can be brought by this is that as users have to express their appreciation of the idea as a regular post in the discussion, convergence starts quite early. We frequently found the initial replies to a new idea to simply be an agreement with it.

Finally, most channels had no mechanism for signaling the end of a discussion, especially through an actionable outcome (e.g. a specification of the new feature, or perhaps a description of why that can't be accomplished) (**CT-3**). This means that many discussions never officially received any conclusion, resulting in confusion about status ("*Is there anything happening to this request?*", GIMP) unnecessary extra discussions ("*Discussed many times before but never resolved*", Audacity), or lack of support for further development ("*One way for a non-programmer to increase the chance that someone will pick this up is to make it as easy as possible. For example creating a wiki page that describes how you would like such integration to work (possibly with screenshots and sketches)*", Moodle). This can result in an environment which does not advance some features because, among other issues, "*responsibilities are unclear*" (P1). The exception to this comes in the CVS and Issue Tracker communications, which feature status tags for a discussion, though as will be discussed in 5.6 (group diversity), these channels come with their own drawbacks.

Bottlenecks in shared materials

Many of the environments evaluated here had limited support for content other than text (**SM-1**). The available support usually comes in the form of attaching some documents, which then need to be downloaded before they can be used. This is a problem as many times users attempt to describe ideas that are better explained through images, video, or a mockup. The initial reaction may be to simply describe it textually, which may lead to confusion, leading to several messages similar to: "*provide a screenshot of case 1 and 2, […] to see if I can understand what you are getting at*" (Audacity) or "*It feels like this is a lot of text for an idea that might better be explained with a diagram*" (Moodle). On the other hand, others may resort to laborious forms of representations. For example, an Audacity user tried to explain his idea by drawing waveforms with text characters. Therefore, having no support for other types of materials can lead users to use laborious ways to better communicate their ideas.

This issue is aggravated by the fact that the materials are disconnected from the discussions, and usually cannot be interacted with, which made communication and collaboration somewhat difficult (**SM-2**). For example, even when someone shared a different form of material (e.g. a mockup), responses were textual (e.g. another user asking for changes to be made). Such changes were seldom integrated into the materials. Furthermore, their lack of interactivity could trigger requests for clarification ("*How does the example you showed work?*", FreeCAD) or for the original user to post screenshots of several scenarios ("*Could you please also post a screenshot with all of these at the same time (as a worst-case scenario)?*", Moodle). In a context where user time is significantly restricted, extra work such as this may hinder further contributions.

Bottlenecks in shared understanding

Perhaps one of the most ubiquitous issues in shared understanding was of a lack of connection and discoverability of parallel or relevant discussions (**SU-1**). We frequently observed users pointing out similar discussions that had already happened in the same channel ("*I suggest to run a forum search for "fgrun" and "t5" to learn more about the latest developments*", FlightGear) or in other channels (FlightGear user posts a link to the issue tracker saying "*that's a long standing (ie 2011) request*"). Consequently, users often missed relevant information or had to look for it. This also came out as an annoyance to users, who had to search for such information. For example, a FlightGear user mentioned that "*having to go to google all the time to search is somewhat annoying so I don't do it all the time.*" (FlightGear), and even when he would find information, it would be "*outdated*", and therefore "*more confusing and misleading (sic) than helpful*" (FlightGear). This lack of discoverability can also lead to redundant discussions or work ("*This is duplicating the work that Bill has done already, isn't it?*", GIMP) or "*this suggestion looks indeed very similar to what was discussed several times on the gimp-developers mailing list*" (GIMP). This issue is reduced in issue trackers, which often feature the functionality of tracking duplicates. It is not, however, common in discussion boards or mailing lists.

The other issue in shared understanding was that the style of conversations generally afforded by these channels resulted in long textual interactions (**SU-2**). As time is a rare commodity in this context, the result is that many people would not read the existing posts. In fact, we commonly find utterances such as "*I only skimmed the first really long ticket*" (Moodle) or "*I also did not read all the old posts here*" (FreeCAD). Sometimes, users voluntarily intervene to improve the situation: "*I read through the long thread and would like to summarize a bit*" (LibreOffice). This, however, was not a common occurrence. And even when it done, sometimes could also lead to confusion: "*The summary could be more clear*" (Drupal).

Bottlenecks in the collaborative, iterative process

The first issue we found related to the collaborative and iterative process is that of lack of versioning of shared materials. As a result, user must manually notify others of changes (**CI-1**). This is not much of an issue in two environments: issue trackers and Wikis. In GitHub, for example, a discussion around a pull request includes information about new commits in between discussion posts, meaning that users can already see. In Wikis, which Moodle participants often used to write specs, there is also a tracking of changes. Nonetheless, users still had to notify others of changes they've made (e.g. "*By the way, I've made a bunch of updates to the spec over the weekend*"). However, failing to correctly update the material or notify others could cause confusion: "*I've just realized that I didn't post the rest of the Instructor Specification. […] Sorry for the confusion.*"

The other issue we identified is the lack of integration with synchronous channels of communication (**CI-2**). Many FOSS projects use channels such as IRC, and other distributed groups often communicate through technologies such as Slack. Throughout discussions, we find messages such as "*Can we discuss this on IRC please, I have quite some comments"* (GIMP), or "*If you would like to discuss this with me, please ping me and we can arrage (sic) a time to talk on Skype, Google hangout, or something*" (Moodle). Synchronous channels offer clear benefits in a situation of quick interactions. The issue is that since this collaboration occurs outside of the scope of the asynchronous communication channels, and therefore their results need it be somehow relayed back to them, otherwise alienating those who use primarily those channels to collaborate. For example, we found instances in which Moodle users note that an announcement was made "*in dev chat and discussed it in the recent developers meeting*" or they "*mentioned on the dev chat (...)*". This involves extra effort in a context in which contributors suffer from lack of resources to contribute as much as they probably would like to, and users who do not frequent those channels will likely miss the information.

Bottlenecks in group diversity

We found one issue related to group diversity: the ubiquitous use of overly technical language and system features. Naturally, technical language is necessary for implementation discussions. However, many times discussions were not at the point of implementation. For example, in a post discussing a new Zotero feature, users express their lack of knowledge to follow the discussion: "*NOT being a programmer, much of the above is 'greek' to me*". In other times, technical language was used in relation to the tools used by the community. For example, users could be referred to look at commit messages to get a status update on the feature ("*If you follow the commits [...] you can* *clearly notice improvements*", Moodle). In fact, interviewee P1 explained that Zotero keeps "*far-reaching*" discussions in the forums (as opposed to GitHub) because GitHub is still perceived as a "*coder's place*", and that people "*who have never coded anything*" may "*feel more comfortable*" in the forums. Such a distinction was also seen in GIMP ("*Let's keep the bug reports focused on technical issues*").

However, this is not exclusive to software-related language, but extensive use of a project's jargon can also pose an issue. For example, in the middle of a discussion about changing outcomes in Moodle, a user intervenes: "*I freely admit that much of the Moodle-speak in what I have looked at here goes straight over my head and that might be why others tend not to comment*", claiming to now provide them with the "*educational practitioner viewpoint*". Therefore, centering interactions around technical aspects (software or otherwise) could draw away those with valuable input, but perhaps less technical knowledge.

Discussion

I summarize the main results as follows:

- The affordances of the channels favor discussion over exploration. This results in new ideas getting lost, and the group focusing on somewhat endless discussions;
- Participants tend to collaborate using text, which is not always the most appropriate method and may lead to issues in shared understanding. Other than code, there is a lack of support for other types of media;
- Communications happen across several different channels, and usually through long blocks of text. Important information for discussions tends to be missed;

- There is a lack of support for iteration over ideas or materials, as well as for synchronous forms of collaboration. This leads to users having to update others of progress manually;
- CVS and issue trackers provide better support for some heuristics, but their developer-centric characteristics may discourage non-developers from participating or may focus discussions too much on code.

These findings also point to important contextual matters. Unlike many CSTs in the literature—designed for supporting groups who come together with the intent of brainstorming—members in the FOSS communities we evaluated rarely do that. Instead, discussions revolve around a very specific (and often detailed) idea proposed by one member, who suggested it motivated by his or her own use case. The issue, however, is that the affordances of these channels (see the divergent and convergent thinking bottlenecks) favor long-winding discussion rather than exploration. If divergence is limited, the community's potential for innovation may be damaged. This points to the need of investigating ways of fostering divergence (as well as better convergence) in a context outside of traditional brainstorming. The results from our heuristic analysis point to ways in which this could be improved. such as differentiating between posts in which an alternative idea is suggested to those in which users discuss the alternatives themselves.

Another interesting finding is related to the use of modern CVS tools, especially GitHub. These tools address, to a degree, many of the issues that older and more general tools have. For example, they can track related or duplicate issues, are integrated with the code that is being discussed, and allow the community to keep track of which discussions are still open or which ones have been resolved. Curiously, those projects that employed such channels tended to have fewer issues coded (see Table 2). However, these results on group diversity show that such environments may still be perceived as hostile to non-developers. Furthermore, these environments seem to afford more technically grounded discussion rather than exploratory ones, making discussions tend towards convergence rather than divergence. Therefore, a viable pathway for better supporting creativity of such groups could be to adapt these environments to be more approachable by non-technical users, while also addressing the issues they share with discussion boards and mailing lists (e.g. separating ideas from discussions about those ideas).

Limitations

The goal for this chapter was to identify problematic elements in the channels' interfaces, as well as observable breakdowns in collaboration, discussing their possible relationship. There are, however, limitations with the methods employed here. They do not allow us to determine causality with certainty. It is unclear how much of the observed issues are in fact due to the affordances of the channels as opposed to some other factor (e.g. community culture or participation levels), especially in a context such as that of FOSS projects, in which participants generally act on heavy time constraints. Future research could probe deeper into this relationship, as well as into the mechanics of their effects on creative output.

Another limitation with the methods employed here is that a heuristic analysis is unlikely to capture all problems with the software (Nielsen & Molich, 1990). I employed three analysts, which is considered to be the baseline number (Nielsen & Molich, 1990).

Adding more analysts would increase the chance of capturing further mistakes, but even so the method would likely not capture all issues. Consequently, the list of issues is not comprehensive, but should rather be considered a list of, perhaps, some of the most salient issues in those channels. Furthermore, an analysis of the content cannot reveal bottlenecks that never even made it to the channels themselves—such as a user that is too shy to post.

Further, there may be question regarding the trustworthiness of the coding scheme. The coding for the discussion content was done in accordance to the general inductive approach detailed by Thomas (2006). Specifically, I employed the "independent parallel coding" strategy for ensuring trustworthiness of the scheme, in which the final set of categories was developed by two independent coders. Each coder worked independently and met on two occasions to discuss and merge categories. Therefore, the scheme is not a product of my individual assessment, mitigating my own biases and reinforcing the trustworthiness of the results. Nonetheless, no metric for agreement was computed between the raters. This does not invalidate the results, given the methodology employed. However, it may open questions on the extent of the reproducibility of the results.

There are also limitations with the interviews. User interviews could not only reinforce the findings from the other two sources, but also expand them, as they would provide insight into the users' thoughts rather than exclusively their output. However, I was unable to recruit a significant number of community members, and therefore not much input from the users was collected. Furthermore, their analysis was only informal and meant to exemplify user expressions of the bottlenecks identified through the heuristic and content analysis. Therefore, the interviews presented in this chapter are used mostly at an exemplary capacity.

In conclusion, given the negative focus of this analysis (on bottlenecks rather than enablers of creativity) I must emphasize that the issues discussed above do not mean that these communities are failing to be creative, nor that the channels they use are poorly designed. Both in the visible outcomes of their collaboration as well as in conversations with their members, we find compelling evidence that creativity is well and thriving, and that users perceive the tools to be fulfilling their needs. The issues described above simply demonstrate that there is room for improvement in the tools they use if they are to better support distributed creativity, as well as pointing to some of the affordances that could be improved. This should by no means be a radical departure from current tools, as they serve an important purpose of a low-barrier of entry to newcomers into the project (such as mentioned by P1), although the exact ways in which these channels could be improved fall outside of the scope of this chapter.

Conclusion

Distributed groups show great potential for creativity and innovation, but the channels they use to communicate may hinder to a degree the creativity of those collaborations, and consequently their potential for innovation. To understand in which ways this may happen, I surveyed the literature on CSTs to generate a list of six heuristics that can be used to evaluate the fitness of collaboration channels for creativity. Using these heuristics as guides, I (with the help of other analysts) performed three kinds of analysis: 1) a heuristic analysis of the communication channels; 2) a content analysis of discussions within those channels; and 3) interviews with FOSS contributors. This surfaced a list of twelve design problems in these different channels. Within these design problems, our content analysis as well as interviews revealed a set of fifteen possible issues for the collaboration within channels that employ those designs. While further research is necessary to better understand the relationships between these design elements and their implications, as well as the degree to which the creativity is affected, these results begin to expose the limitations common collaboration channels may have on the creativity of distributed groups.

Before further considering how to improve the creativity of existing online groups (discussed in chapter 6), I now turn the focus of this dissertation towards the approaches for enhancing crowd brainstorming described in the literature. Much of this work does not happen in the context of real online communities such as those evaluated here, but rather in crowds such as Amazon's Mechanical Turk platform. Nonetheless, they describe promising ways of supporting large scale creativity. Consequently, advancing these techniques can eventually lead back to improvements for existing communities. Therefore, in the next two chapters, I focus on existing research techniques for supporting crowd creativity and how they could be improved through increased attention (chapter 4) and adaptation (chapter 5).

CHAPTER 4

IMPROVING INSPIRATION EFFECT THROUGH MICROTASKS

With the advent of crowdsourcing, people can now collectively accomplish a wide range of tasks that could not otherwise be done by a single human or computer. One approach to crowdsourcing that has stood out is the use of micro-task markets such as Amazon's Mechanical Turk (MTurk) (Kittur, Chi, & Suh, 2008). In this approach, many workers perform small tasks that together approximate the quality of experts. Using micro-task markets, researchers have been able to achieve good results on a wide variety of tasks (Bernstein et al., 2010; Chilton et al., 2013). In this chapter, I leverage a similar micro-tasks paradigm to increase the diversity of idea generation.

Creativity thrives on diversity and exploration. It is about creating something that is both novel, breaking away from common knowledge or practices, but at the same time being appropriate or useful (Hennessey & Amabile, 2010). From designing T-shirts (www.threadless.com) to solving tough technical challenges (www.innocentive.com), there are many examples of the crowd performing tasks that rely on their creativity.

A great number of people will generate a great number of ideas. Furthermore, the heterogeneity of the crowd can increase the potential of ideas being sparked that otherwise wouldn't (Dennis & Williams, 2003). However, there are also issues that need to be carefully considered in a system that tries to tap into the crowd's creativity. Issues such as cognitive interference or social loafing can increase together with the number of ideators (Dennis & Williams, 2003). Therefore, crowd ideation needs to be carefully designed in order to improve, not hinder the creative output.

A popular method used for generating ideas is typically brainstorming, which seeks to increase the number of ideas generated by encouraging intensive exploration of ideas while restricting criticism (Osborn, 1963). In the crowd context, just like in smaller groups, people have tried to enhance idea generation during brainstorming sessions in different ways, many times employing other individuals or workers, outside of ideation, to do tasks whose output will benefit ideators. I call these tasks *peripheral tasks*. The result of their work is then presented in some way to crowd ideators. For example: Yu et al. (2014a) had workers generate problem schemas, and subsequently used them to enhance ideation performance of other workers in a subsequent study. However, the extra cognitive effort that is required to perform these tasks could potentially benefit ideators as much as just using their results.

This chapter, therefore, examines the effect that performing peripheral tasks has on ideation. More specifically, it embeds three types of peripheral tasks—rating, similarity, and combination—into an online brainstorming session. I explore the following questions:

1. How does performing peripheral micro-tasks affect ideation performance?
2. Do different types of peripheral tasks affect ideation differently? If so, how?

Exploration of these questions could allow ideation systems to move from passive to active forms of inspiration and support, resulting in more data collection during an ideation session, aiding in convergent tasks such as idea selection. A similar approach has been explored by Siangliulue and colleagues through the IdeaHound system (Siangliulue et al., 2016). IdeaHound allows users to physically cluster semantically related ideas together in a virtual workspace. This organization enables the system to infer a semantic model of the ideas. My approach differs in that it makes this data collection explicit rather than implicitly building it in the UI interactions of the system. In other words, rather than inferring semantic relatedness by examining how ideators cluster ideas together, explicitly asking them to judge the similarity of two ideas. The focus, however, is on how doing these tasks affects ideation performance, rather than examining their result. This chapter contributes to my overarching goal of supporting large-scale creativity by testing a mechanism that can increase the efficacy of large-scale ideation support strategies.

In the remainder of this chapter, I describe a simple system built to allow ideators to perform small tasks during ideation, and describe metrics for its evaluation, including a tree-based representation of individual ideators' performance. I then describe four iterative experiments, evaluating how their combined results answer the questions above. Generally, I find that tasks are just as useful as simple idea exposure, with rating and combination tasks even outperforming it in certain situations. I also explore how inspiration size, frequency, and homogeneity affects ideation.

The work presented in this chapter is based on Girotto et al. (2017).

The System

I developed an online ideation system that enables the creation of timed asynchronous ideation sessions. It also has a mechanism for seeing other people's ideas upon request via an inspiration button, thus allowing ideators to *pull* inspirations whenever they choose to do so. This pull approach is in line with existing literature (e.g.

Chan et al., 2016b; Siangliulue, Chan, et al., 2015). An alternative to this approach would be to *push* inspirations at regular time intervals. This would ensure that every ideator was exposed to the same number of inspirations, allowing a clearer comparison of the effect of the different types of tasks. However, one of my goals was to see if embedding tasks into inspirations would detract from users' interest in using the inspiration mechanism or decrease performance. A push approach would hinder me from exploring this. Furthermore, the SIAM model predicts that pushing inspirations could negatively affect performance (Nijstad et al., 2003), since it could interrupt users' train of thought. In fact, Siangliulue et al. found issues with fluency using a push approach (Siangliulue, Chan, et al., 2015). Therefore, I allowed users to request inspirations on demand.

The system is comprised of four main parts (Figure 1). Although the figure depicts the system in its final iteration, its overall structure as described in this section was maintained throughout the sessions, with a few incremental differences that will be pointed out in each experiment's section. At the top (A), the system displays instructions, the problem definition, and a timer. On the left is the ideation panel. It consists of a form for entering an idea along with a list of user defined tags associated with it (B), and a list of the user's previously submitted ideas and tags (C). On the right side is the inspiration panel (D). When the button is clicked, the user is presented with a set of ideas and depending on their condition, a task associated with them. This mechanism draws randomly from a pool of ideas generated in previous experiments.

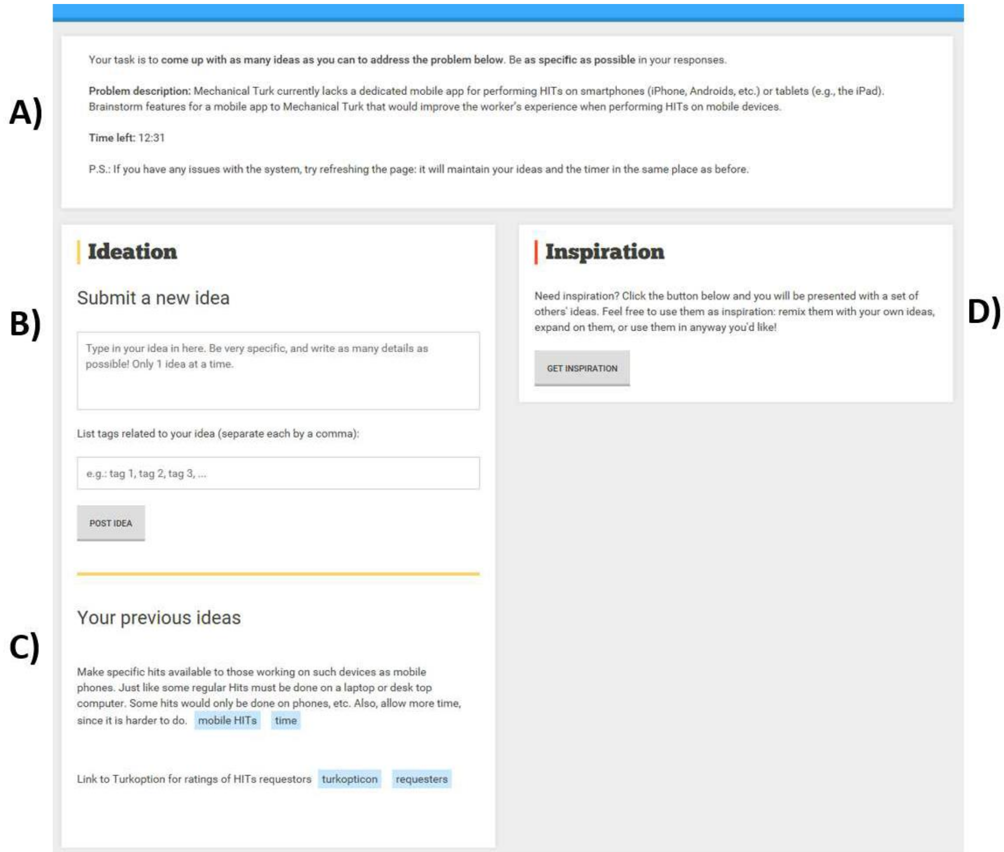

Figure 1. Screenshot of the ideation system as used in the final experiment. It is comprised of the following parts: A) problem description and timer; B) idea submission input; C) list of the users' submitted ideas; D) inspiration panel.

When users access the system, they first see a page stating how much time the session lasts and asking them to move forward only if absolutely sure that they can commit their full attention for the specified amount of time. Following that, users would see another page describing the system, including the inspiration mechanism (if any), and how to use it. Upon finishing the instructions, users begin the ideation session. After the timer is done, the system presents users with a thank you message, a user ID (used for payment), and a link to a short post-session survey.

For every study in this chapter, the problem that ideators were tasked to ideate on was: "*Mechanical Turk currently lacks a dedicated mobile app for performing HITs on smartphones (iPhone, Androids, etc.) or tablets (e.g., the iPad). Brainstorm N features for a mobile app to Mechanical Turk that would improve the worker's experience when performing HITs on mobile devices. Be as specific as possible in your responses.*" This task, suggested by Krynicki (2014), was chosen because it has been successfully used in previous studies (Chan et al., 2016a; Krynicki, 2014) and MTurk users have knowledge about the issue and may be motivated to contribute to it, as it could increase their opportunities for engaging with HITs and improving their income. Both motivation and knowledge are key to creativity (T. Amabile, 1983).

Metrics

For each study, I report the following metrics:

- *Fluency*: number of ideas generated by the user.
- *Number of inspirations*: number of times the user clicked the inspiration button.
- *Inspiration influence*: a user's average similarity between an idea and the most similar of its preceding inspirations.

More central to my interests, however, are metrics of *breadth* and *depth*, which I extracted from an *ideation tree*, described below. Tree representations have been previously suggested to measure or visualize ideation outcome (Ivanov & Cyr, 2006; Nelson, Wilson, Rosen, & Yen, 2009), and the semantics of the different branches of a tree can reflect the usual discrete categorization of ideas traditionally used in creativity research (Plucker & Makel, 2010), while their depth can represent the notion of ideation within one category (Nijstad & Stroebe, 2006). This tree is built from a chronological list of user actions—they either add a new idea or request an inspiration. In the tree, similarity between ideas is measured using Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998). In this chapter, the LSA corpus was built on 5640 ideas generated to solve the same problem that I explore in this chapter. This corpus comes both from my own pilot studies (2115 ideas) and the corpus shared by Chan et al. (2016b) (3525 ideas). Figure 2 shows the tree and idea pool in five different points in time during a user's ideation. The tree building algorithm works as follows:

1. We add the first user idea as the child of a dummy node;
2. For the second user generated idea, we compare it either with every node that is already in the tree, or with every inspiration previously seen. If the LSA similarity to any of those is greater than a given threshold (we used 0.5), we add it as a child of the most similar node. In this point in time, idea 2 was most similar to idea 1, and is added as its child;
3. At the third point in time, the user has generated a third idea. Again, we compare it to every node already in the tree. In this case, none of the similarities exceeded the threshold, so this idea is added as a new child of the root node, representing an estimated new category of ideation. Also note that between t2 and t3 the ideator has requested an inspiration, which is added to the idea pool but not to the tree;
4. The user's 4th idea is compared to every node in the idea pool. In this case, a previously seen inspiration is the most similar. Therefore, we add it as a child of the inspiration node, which we then add as a new child of the root node;

5. Finally, another idea has been generated by the user. It is compared to previous ideas and inspirations and is added as a child to the most similar node, which is the first one.

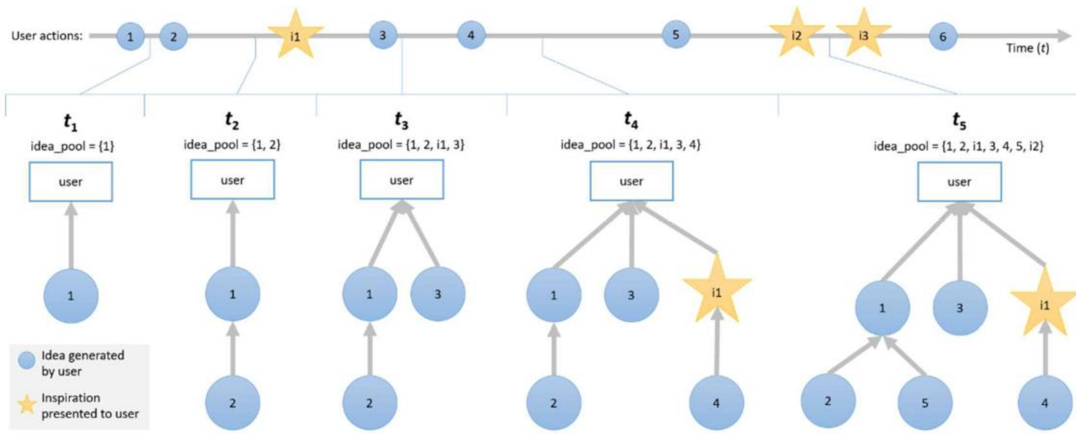

Figure 2. Snapshots of the ideation tree and idea pool in five different points of time.

From this tree, I extract the two metrics:

- *Breadth*: the number of children in the root node. These were the ideas that, at the time they were added, were not similar enough to be considered a continuation of another idea, therefore creating a new branch of ideas. For example, in Figure 2 at t5, the breadth would be 3.
- *Depth*: the number of nodes in the branch with the most number of nodes. For example, in Figure 2 at t5, the depth would be 3.

As a check on this measure, I additionally calculated the metrics described by Chan et al. (2016b), also built using LSA. In Chan et al. (2016b), breadth was the mean distance between each pair of ideas generated by a user. Depth was the maximum similarity between the ideas generated by a user. My metrics are significantly correlated with these, at $r = 0.650$, $p < 0.001$ for breadth, and $r = 0.564$, $p = 0.001$ for depth. And while there may be concerns relative to the metric's sensitivity to the threshold value, I have found that threshold changes in either direction do not result in drastic changes in the results. For example, changing it to 0.7 resulted in a mean difference of 2.3 (SD=2.2) in breadth. Therefore, I believe that this metric is both valid and capable of more accurately representing the notions of breadth and depth.

Experiment 1: Rating Task

For the first experiment, I chose to start the exploration with rating tasks. Due to their simplicity, they can easily and quickly be done by any worker, and they are very effective for supporting convergence processes. I used an earlier version of the system than the one described in the system section, which differed as follows: the input box for the idea was on the top panel, along with the problem description; users did not have to input tags for their ideas; Lastly, the inspiration box did not have any instructions regarding how inspirations could be used. This study had three conditions:

1. *Control*: This condition is equivalent to a nominal group in typical brainstorming settings. There is no inspiration panel, and thus no external stimulus. Users type their ideas and can see the list of their own ideas.
2. *Exposure*: In this condition, the inspiration panel is visible. When the inspiration button is clicked, it displays one idea from the pool of past ideas, without any task associated with it. The idea disappears when the user clicks the "done" button.
3. *Rating*: This condition is similar to the exposure condition. However, when the inspiration button is clicked, in addition to the idea, users also received a task

prompting them to rate the inspiration idea in 2 dimensions: originality and practicality (Figure 3). After submitting the rating, the idea disappears.

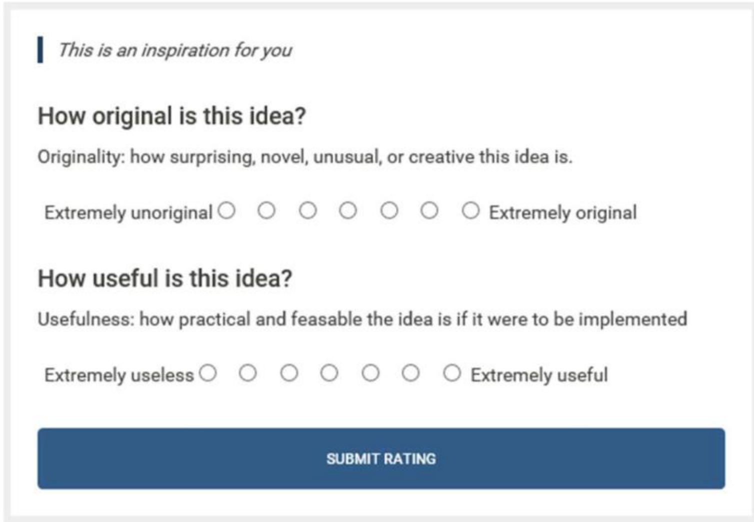

Figure 3. Rating task interface.

I published a MTurk HIT that directed workers to this system. 60 workers participated in this study (at least 1000 completed HITs, approval > 95%, US only), but one (exposure condition) was excluded from the analysis due to an abnormal number of inspirations requested (142). In total, 559 ideas were generated. Each worker ideated for 18 minutes, filled out a small survey at the end of the session, and was compensated $2. Workers also received an ideation qualification on MTurk (awarded after every experiment). Subsequent experiments reported here required workers to not have this qualification, thus ensuring participants were unique for each session.

Tables 1 and 2 summarize the metrics across the different conditions. A one-way ANOVA test shows no significant difference in fluency, $F(2,56) = 0.713$, $p = 0.495$. There was a marginally significant difference in number of inspirations requested across the two conditions, F(1,38) = 3.855, p = 0.057. There was, however, a difference in inspiration influence between the exposure and rating conditions, F(1,38) = 9.855, p = 0.003.

Table 1

Fluency and inspiration metrics for experiment 1

Condition	Workers	Ideas/worker	Insp. / Worker
Baseline	19	10.37 (4.16)	-
Exposure	19	9.53 (4.62)	12.16 (10.64)
Rating	21	8.62 (5.02)	6.67 (6.80)

Table 2

Breadth, depth, and influence for experiment 1

Condition	Breadth	Depth	Influence
Baseline	5.37 (2.54)	4.74 (3.69)	-
Exposure	6.40 (3.20)	3.25 (1.74)	0.23 (0.90)
Rating	5.10 (2.70)	3.29 (1.52)	0.12 (0.12)

I calculated a Mixed Generalized Linear Model (GLM) with breadth as outcome variable, condition as factor, and the fluency as covariate. I included the interaction between condition and fluency in the model. I found a marginally significant interaction effect between condition and fluency, F(2, 56) = 3.078, p = 0.054, and no main effect of condition on breadth, F(2,56) = 1.374, p = 0.262.

As the depth of user ideas followed a negative binomial distribution rather than a normal distribution, I conducted a negative binomial regression with depth as outcome, condition as factor, and fluency as covariate. The interaction between fluency and condition was included in the model. I found no significant interaction effect or main effect of condition on depth, Wald Chi-Square = 4.099, p = 0.129.

Discussion for Study 1

While we see only a marginal effect of the interaction between condition and fluency on breadth, there is no clear advantage in any condition. The fact that breadth seemed to be more affected than depth may spring from inspirations being randomly drawn from the pool of ideas, which will likely create a heterogeneous set of examples. Past work has shown that a heterogeneous set of examples will improve diversity of ideas(Nijstad, Stroebe, & Lodewijkx, 2002; Siangliulue, Arnold, et al., 2015). Having no clear advantage could mean a problem either in the intervention (e.g. it is too simple) or in how users performed it (e.g. they did not attend to it). There may also have been confusion on how users should use the ideas in the rating task. For example, a user declared feeling that the inspiration they got would invalidate using that idea: "*I think it hindered me more than it helped because it just provided an example that I then couldn't use*". Perhaps guidelines might be effective in helping users better use the inspirations.

Experiment 2a & 2b: Similarity Choice Task

In experiment 1, there was no clear advantages over the baseline, even though there was a larger influence in the exposure condition. Given these results, I decided to change the task to similarity comparison, the number of ideas displayed, and to add a clarification on how they could use the inspirations (see the text at the top of Figure 4).

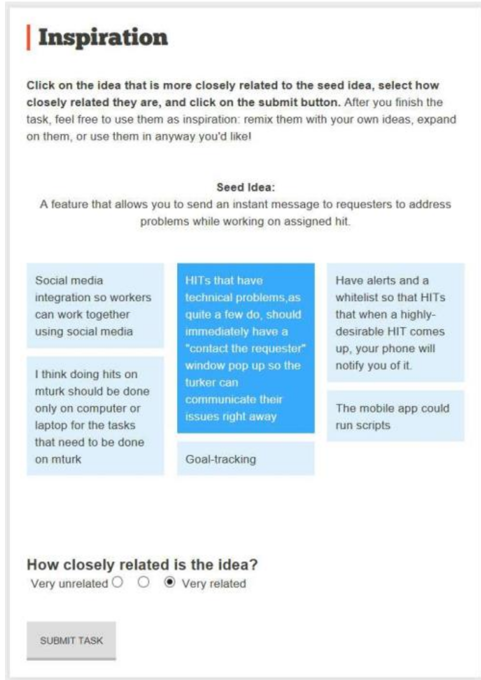

Figure 4. The task panel for condition 3. Users were shown a seed idea along with 6 other ideas and were asked to click on the most similar one (in this case, the user clicked on the dark blue idea), as well as rating their degree of relationship.

Experiment 2A

In this experiment, the task condition presents the user with one seed idea along with 6 other ideas, asking him or her to choose the most similar to the seed (see Figure 4). The number of ideas was chosen to maximize the possibility of similar ideas being shown, as well as to explore the result of a more dramatic increase in the number of ideas shown per inspiration. I expected this to yield a stronger influence on ideators' breadth, as they would be exposed to more ideas. I also hypothesized that similarity comparisons would force to user to think more abstractly about the ideas in order to find common features between them, thus reducing fixation and possibly improving breadth.

This second experiment followed the same method as the first, with the two key differences above. 60 workers participated in this study (at least 1000 completed HITs, approval > 95%, US only). In total, 492 ideas were generated. Each worker ideated for 18 minutes, filled out a small survey at the end of the session, and was paid $2.

Table 3

Fluency and inspiration metrics for experiment 2a.

Condition	Workers	Ideas / Worker	Insp. / Worker
Baseline	20	7.45 (5.51)	-
Exposure	22	8.50 (4.34)	2.00 (2.19)
Similarity	18	8.61 (2.87)	2.28 (1.77)

Tables 3 and 4 summarize the metrics for this experiment. A one-way ANOVA test shows no difference in fluency, $F(2,57) = 0.416$, $p = 0.661$, or number of inspirations requested between the exposure and task conditions, $F(1,28) = 0.261$, $p = 0.612$. Finally, similarly to the last study, there is a difference in inspiration influence. This time, however, the task condition displayed a higher influence than the exposure, $F(1,28) = 4.59$, $p = 0.039$ (see Table 4).

Table 4

Breadth, depth, and influence for experiment 2a

Condition	Breadth	Depth	Influence
Baseline	4.90 (3.43)	2.40 (1.31)	-
Exposure	6.36 (2.95)	2.05 (0.785)	0.11 (0.03)
Similarity	5.78 (2.15)	2.28 (0.82)	0.13 (0.03)

I calculated a Mixed GLM with breadth as outcome variable, condition as factor, and fluency as covariate, finding no significant difference between conditions, $F(2,57) = 1.962$, $p = 0.150$. As in the last study, I conducted a negative binomial regression for depth. With condition as factor and number of ideas as covariate, there was no significant difference, Wald Chi-Square = 2.108, $p = 0.348$.

Discussion for Experiment 2A.

Unlike the first study, the task condition yielded a significantly higher influence than the exposure condition, but this did not translate into an improvement in ideation breadth or depth. In general, all three conditions appeared to be very similar with respect to breadth and depth, despite the small but significant difference in influence. This is not so surprising when you consider the low number of inspirations requested for both inspiration conditions—close to 2. It is likely that the great number of ideas per inspiration either overwhelmed users or provided them with what they judged to be enough inspiration for a long stretch of time.

Experiment 2B: smaller inspirations, controlled pool.

In experiment 2A, there was no meaningful difference across conditions, likely due to the very small number of inspirations requested in both experimental conditions. Therefore, I reduced the number of ideas per inspiration to 3. I also controlled the pool of ideas. I went through the existing idea pool and generated 40 different groups of 3 ideas. The goal was to create sets of ideas that shared similar elements, making the choice task easier, while at the same time having different features. For example, the idea "*Notifications such as sound or vibration when a new hit is available*" was grouped with "*sounds effects so people know when to do the surveys and also tools to see how well they are doing*" and "*The app would have alarms, bells, or sounds to notify of particular work* *or requesters*”. The three ideas in the group would always come together, but in random order.

The method for this experiment changed in one significant way: I increased ideation time from 18 to 25 minutes, to collect more data on how ideation changes as fluency increases. I also increased the target number of users per condition (see Table 5). 89 workers participated in this study (at least 1000 completed HITs, approval > 98%, US only. In total, 863 ideas were generated. Workers also filled out a small survey at the end. Each worker was paid $3.50.

Table 5

Fluency and inspiration metrics for experiment 2b

Condition	Workers	Ideas / Worker	Insp. / Worker
Baseline	35	8.43 (4.74)	-
Exposure	27	9.53 (5.95)	12.59 (12.29)
Similarity	27	9.15 (5.11)	5.74 (5.18)

Table 6

Breadth, depth, and influence for experiment 2b

Condition	Breadth	Depth	Influence
Baseline	5.31 (2.99)	3.17 (2.62)	-
Exposure	8.07 (3.79)	3.33 (1.68)	0.14 (0.04)
Similarity	6.33 (3.75)	2.78 (1.45)	0.16 (0.04)

Tables 5 and 6 summarize the metrics for this experiment. A one-way ANOVA shows a significant difference in fluency, $F(2,86) = 3.528$, $p = 0.034$. A post hoc Tukey test shows a significant difference between baseline and exposure conditions, $p = 0.031$, but no difference between baseline and task ($p = 0.854$) or exposure and task ($p = 0.139$). There was also a significant difference in number of inspirations between the exposure and similarity conditions, $F(1,52) = 7.119$, $p = 0.01$. However, this time no significant differences were found in inspiration influence, $F(1,47)$ 2.019, $p = 0.162$.

I calculated a Mixed GLM with breadth as outcome, condition as factor, and fluency as covariate, including the interaction between condition and fluency. There was a marginally significant interaction between condition and fluency, $F(2,83) = 2.88$, $p = 0.062$, but no main effect of condition on breadth, $F(2,83) = 1.269$, $p = 0.286$.

For depth, a negative binomial regression with condition as factor, fluency as covariate, and including the interaction between fluency and condition found a significant interaction between fluency and condition, Wald Chi-Square = 10.003, $p = 0.007$, but no significant main effect of condition, Wald Chi-Square = 4.550, $p = 0.103$. A pairwise comparison shows a difference only for high fluency ideators (1 SD above the mean). In this case, those in the control condition ($M=6.31$, $SE=0.850$) performed significantly above both exposure ($M=3.84$, $SE=0.425$, $p = 0.009$) and similarity ($M=3.35$, $SE=0.584$, $p = 0.004$) conditions.

I also divided both halves of the ideation and analyzed their breadth and depth separately. This was done since the effect of inspirations on users is likely not constant across the session, as they will likely be able to generate more ideas by themselves at the beginning of the session than at the end, when inspirations may be more useful. Thus, looking at the metrics over the entire session may wash out some effects.

A Mixed GLM with breadth for the first and second halves (ran separately) as outcome variables, with condition as factor, and fluency as covariate yielded no main effect of condition on the first half breadth, $F(2,85) =2.704$, $p = 0.073$. On the second half, however, it yielded a main condition effect, $F(2,83) = 3.527$, $p = 0.034$, as well as a significant interaction on condition and fluency, $F(2,83) = 6.957$, $p = 0.03$. In pairwise comparisons, a difference was seen for low fluency ideators (1 SD below the mean), where the control condition (M=1.54, SE=0.322) was significantly superior to the task condition (M=0.38, SE=0.381, $p = 0.022$), but was not significantly different than the exposure condition (M=1.06, SE=0.447, $p=0.386$).

For the first half depth metric, a negative binomial regression with condition as factor, fluency as covariate, and including the interaction between condition and fluency yielded a significant main effect, Wald Chi-Square = 6.48, $p = 0.039$, and a significant interaction between condition and number of ideas, Wald Chi-Square = 7.46, $p = 0.024$. For low fluency ideators (1 SD below the mean), the exposure condition (M=1.99, SE=0.348) significantly outperforms the control condition (M=1.05, SE=0.22, $p = 0.042$), but it was not significantly different to the task condition (M=1.74, SE=0.409, $p=0.638$). No pairwise differences were seen for high fluency ideators. The second half presented no significant interaction or main condition effect, Wald Chi-Square = 3.362, $p = 0.186$.

Discussion for Study 2B.

To summarize this study, there was a significant difference in fluency only between the exposure condition over control. The exposure condition also saw more inspiration requests. Baseline high fluency ideators outperformed the others in overall depth, low fluency baseline outperformed task in 2nd half breadth, and low fluency exposure outperformed baseline in 1st half depth. In other words, the inspirations not only did not help, but actually hindered the depth of high fluency ideators. It is possible that the closely related nature of the inspirations promoted fixation for them, thus detracting from their second half depth. Finally, for low fluency ideators, we find exposure helping them in first half depth, but tasks detracting from their second half breadth.

Experiment 3: Comparison Across Task Types

I conducted a final study in order to compare the two previous task types as well as a new one: combination. Combination tasks involve not only convergent processes, but also a divergent one—the generation of the new, combined idea (Kohn et al., 2011). While this can happen naturally during ideation, this task explicitly forces it to happen. Therefore, I expect a positive impact of combination on breadth. I also reverted to a completely random inspiration retrieval. The method remained the same as the one employed in experiment 2B, with the difference being that there are five conditions (control, exposure, 3 task types).

150 workers participated in this study (at least 1000 completed HITs, approval > 98%, US only), but 7 workers were not included in the analysis, as they either wrote unrelated ideas (n=1), generated unrelated tags (e.g. “tags 1”, n=4), or didn’t complete the post session questionnaire (n=2). In total, 1480 ideas were generated. Workers ideated for 25 minutes and filled out a small survey at the end of the session. Each worker was paid $3.50.

Table 7

Fluency and inspiration metrics for experiment 3.

Condition	Workers	Ideas / Worker	Insp. / Worker
Baseline	29	11.38 (7.178)	-
Exposure	28	10.57 (6.143)	7.70 (6.92)
Rating	27	8.48 (4.136)	4.28 (4.86)
Similarity	31	11.52 (6.45)	8.77 (5.36)
Combine	28	9.57 (5.647)	3.16 (2.51)

Tables 7 and 8 summarize the metrics for this experiment. A one-way ANOVA shows no significant difference in fluency across conditions, $F(4,142) = 1.276$, $p = 0.283$. A one-way ANOVA for number of inspirations between conditions shows a significant difference in the number of inspirations requested, $F(3,110) = 8.022$, $p < 0.01$. A post hoc Tukey test shows that the exposure and similarity conditions were significantly higher than the rating and combination conditions ($p < 0.05$), but not from each other. As for influence, a one-way ANOVA shows no significant difference, $F(3,105) = 1.285$, $p = 0.283$.

Table 8

Breadth, depth, and influence for experiment 3.

Condition	Breadth	Depth	Influence
Baseline	7.86 (4.086)	3.28 (2.52)	-
Exposure	8.00 (4.830)	2.61 (1.52)	0.14 (0.05)
Rating	6.48 (3.887)	2.56 (1.76)	0.15 (0.07)
Similarity	8.74 (4.289)	2.58 (1.52)	0.12 (0.04)
Combine	8.07 (4.48)	2.07 (1.15)	0.13 (0.05)

A Mixed GLM with breadth as outcome variable, condition as factor, fluency as covariate, and including the interaction between condition and fluency yielded a significant interaction of condition and number of ideas on breadth, $F(4,133) = 3.736$, $p = 0.006$, but no main effect of condition, $F(4,133) = 1.823$, $p = 0.128$. For average fluency ideators (10 ideas), a pairwise comparison shows a significant difference between the control ($M=7.20$, $SE=0.39$) and combine ($M=8.38$, $SE=0.39$) conditions, $p = 0.037$. There are also significant differences for high fluency ideators (1 SD above the mean, fluency = 16 ideas), in which the control condition was outperformed by all other conditions (pexposure = 0.018, prating = 0.010, psimilarity = 0.046, pcombine < 0.001), but they were not significantly different among themselves. Figure 5 depicts the regression lines for the different conditions.

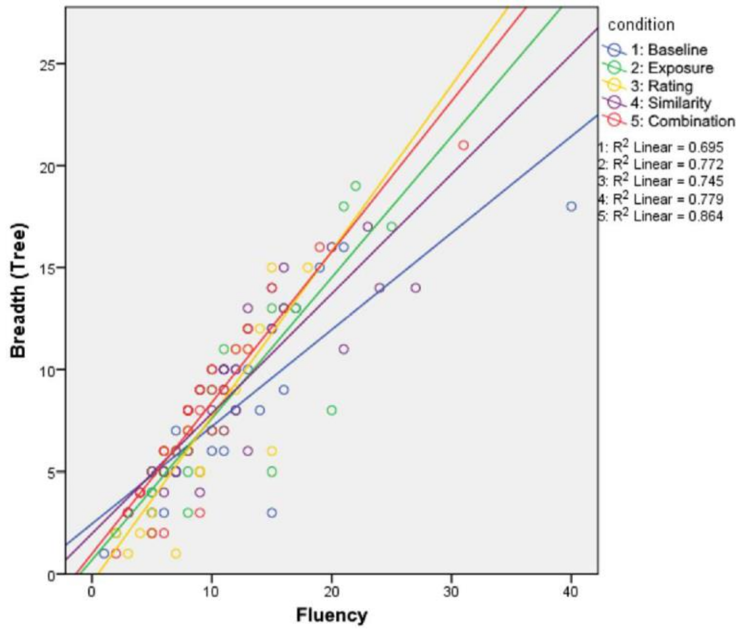

Figure 5. Regression lines for breadth by fluency.

Again, I calculated the breadth metric for each half, and found that a Mixed GLM with first half breadth as outcome variable, condition as factor, and fluency as covariate yielded no significant effect of condition, $F(4,137) = 1.342$, $p = 0.257$. However, using the second half breadth as outcome variable and including an interaction between condition and fluency yields a significant interaction between condition and fluency, $F(4,133) = 7.197$, $p < 0.001$, and a main effect of condition, $F(4,133) = 2.725$, $p = 0.032$. Figure 6 shows the marginal means for second half breadth across the different conditions, with fluency fixed at 1 SD below (4 ideas) and 1 SD above the mean (16 ideas). No significant difference is seen for low fluency ideators. For high fluency ideators, however, we see that that the three task conditions significantly outperformed the baseline ($p < 0.001$). When compared to the exposure condition, however, only the rating and combination conditions significantly outperformed it.

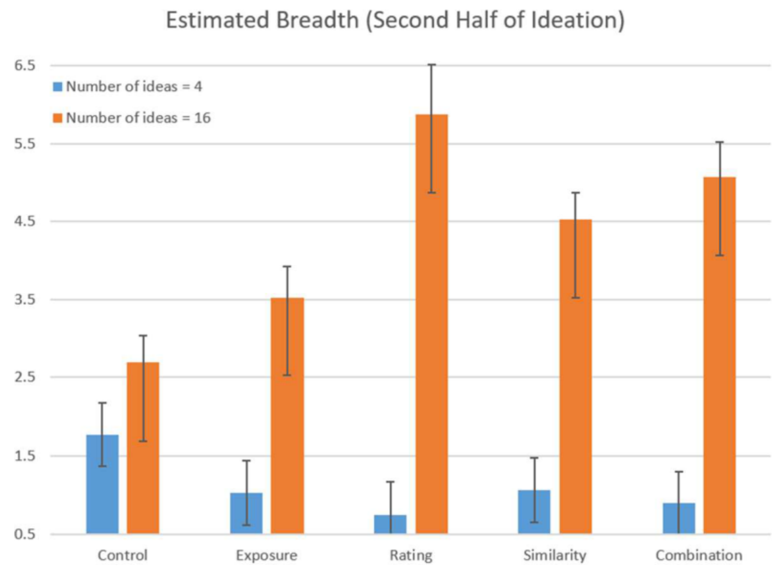

Figure 6. Marginal means and std. error for breadth on the second half of ideation.

For overall depth, a negative binomial regression with depth as outcome variable, condition as factor, and fluency as covariate found no effect of condition, Wald Chi-Square = 5.456, $p = 0.244$. The same model, but separately testing first and second half depth as outcome variables also yielded no significant differences, 1st half Wald Chi-Square = 1.469, $p = 0.832$, 2nd half Wald Chi-Square = 3.422, $p = 0.49$.

Discussion

Through four experiments, I explored the integration of peripheral microtasks as part of an ideation session. Experiment 1 compared the rating task with simple exposure, finding very little influence. Experiments 2A and 2B increased the number of ideas and evaluated similarity tasks, pointing to limitations with quantity and homogeneity of inspiration ideas. Finally, experiment 3 compared all three task types together. I now discuss the main points from the combined results.

Tasks performed as good as exposure, outperforming it in some cases

Experiment 3 shows combination tasks outperforming the baseline for average fluency ideators. As for high fluency ideators, we find all conditions outperforming the baseline. However, when we isolate the second half of ideation, we find significant differences between the types of inspiration. The rating and combination tasks significantly outperformed the exposure condition, while similarity significantly outperformed control. One explanation for this difference is that these tasks were more cognitively demanding than the similarity and exposition inspirations. But unique characteristics of the tasks may explain them further. While usual brainstorming rules discourage criticism of ideas (Osborn, 1963), there is evidence that criticism may foster exploration (Charlan J. Nemeth, Personnaz, Personnaz, & Goncalo, 2004), which could partially account for the better performance of the rating task. Alternatively, it is possible that the rating scales provided users with a structure that guided them in generating ideas or evaluating inspirations. As for the combination task, this result was in line with our expectation, as the task also involves a divergent step (Kohn et al., 2011), which could foster breadth of exploration.

Fewer effective inspirations may be better than many ineffective ones

It is interesting to note that the two most effective conditions had the lowest number of inspiration requests. This may lead to the conclusion that the cognitive load of an inspiration may be more important than the number of times it is used. In other words, fewer but more effective inspirations can be better than having many less effective inspirations. An alternative explanation is that since these users requested fewer inspirations, they had more time to ideate, thus increasing breadth. However, since the fluency was not different across conditions, this is an unlikely explanation.

Inspiration effects depend on timing and fluency

The studies, especially experiment 3, highlight that inspirations may influence different users at different times. On experiment 3, for example, we see significant differences only for average or high fluency ideators. This is not surprising, as low fluency ideators may simply not be engaged enough to attend to the task or the inspirations, regardless of condition. Furthermore, results were mainly seen on the second half of ideation. This is intuitive, since at later points in time ideators are more likely to be running out of ideas (Nijstad & Stroebe, 2006), and thus may be more susceptible to the inspirations. This suggests that a “one size fits all” approach does not work. It may prove useful for crowd ideation support systems to restrain inspirations for a latter phase of ideation, or to initially target fluency improvement.

Very simple or complex inspirations have no effect

Studies 1 and 2A, while exploring two different types of tasks, shed light on lower and upper limits when concerning the number of ideas that can be presented for each inspiration. With both one (experiment 1) and seven (experiment 2A) ideas per inspiration, we see no significant difference between conditions. On the lower end, this lack of effect happened despite a considerable number of inspiration requests. This could be due to the simplicity of the inspirations not fostering attention to the ideas, or to users not knowing how to use the inspirations, as previously discussed. On the higher end, the lack of effect likely happened due to the low number of requests. At the end, we see better effects with inspirations containing three ideas each. This could, however, vary depending on the inspiration type (e.g. a combination task of size 6 could be considerably more demanding than a similarity task of size 6), or even nature of ideas (homogenous idea sets may be less cognitively demanding, allowing more ideas per inspiration).

The homogeneity of idea sets can influence the effects

While most results were seen in breadth, we see a different pattern in experiment 2B, where the inspiration idea sets were manipulated to be more homogenous. In it, exposure outperformed control in first half depth, and control outperformed task in second half breadth. This could be explained partially by the homogenous nature of the ideas, as previously discussed. This indicates that the nature of the inspiration sets is highly influential in the outcome (Siangliulue, Arnold, et al., 2015). Therefore, the effect of different levels of homogeneity and task types should be explored in future work.

Some limitations with this chapter must be noted, the first being the metrics. While the tree-based metric is consistent with previous practices and results, it needs further evaluation. A comparison with similar trees built by human experts would shed light on its performance. Alternatively, graph-based metrics could also be devised in order to better represent the inherent uncertainties in automated textual analysis (e.g. ideas could be linked to more than just one parent idea, with edge weights representing their similarity). Furthermore, we do not explore measures of creativity, whereas past research has used MTurk workers to do that (Chan et al., 2016a; Siangliulue, Chan, et al., 2015; Yu & Nickerson, 2011). However, I have found workers to have very low degrees of agreement among themselves, and therefore I do not report these measures. There are also limitations to the pull approach of inspiration used in this chapter. While it allowed me to compare the performance in a natural setting, the numerical differences in inspiration requests limit my ability to clearly determine the effects of the different tasks. Finally, I do not explore the results of the tasks (e.g. the quality of the ratings). While this exploration is outside of the scope of this chapter, past results are encouraging in the potential of peripheral crowd work to yield useful outcomes such as a semantic model of ideas (Siangliulue et al., 2016).

In conclusion, the work presented here also prompts a discussion on the nature of microtasks. Much of the attractiveness and benefits of microtasks stems from their low barriers of entry (in both time and background knowledge) and low contextual requirements (Kittur et al., 2011). The work presented here does maintain the short completion time of the tasks, taking advantage of it to use these tasks as mid-ideation inspiration. On the other hand, it does break with the assumptions of no background knowledge or contextual awareness. This brings limitations to the usual microtask approaches, as it limits the pool of possible users and independence from individual workers. On the other hand, it may enable results that are not possible without contextual awareness or domain knowledge. Breakdowns due to lack of context have been seen, for example, in Teevan et al. (2016a)'s exploration of a writing workflow. They report users having to talk outside of the system to appropriately infer context and evaluate each other's input. The same may be true for creativity workflows. For example, a microtask workflow for rating the creativity of ideas based on the established consensual assessment technique (Hennessey et al., 2011) may not be feasible without knowledge and context awareness, important requirements for that technique. Therefore, breaking with traditional microtask assumptions can limit the number of workers, but enable new workflows and results.

Conclusion

In this chapter, I have analyzed the effect of performing three different types of tasks normally done by other crowd workers: rating, similarity, and combination. This was done through four subsequent experiments on MTurk to evaluate how they compare to idea exposure or individual ideation. Using breadth and depth metrics based on an ideation tree, I found the performance of task inspirations to be as good or better than simple idea exposure. I also found that the effect of inspirations depends on the fluency of ideators and the period in which it is used. Finally, I found evidence that the homogeneity of inspirations influences the outcome. Therefore, this chapter provides some support and guidance in explicitly embedding microtasks into ideation, which will not only be generating information useful for convergent processes but will also aid ideators in improving the divergence of their idea generation.

Nonetheless, the approach taken in this chapter still relies on a randomized selection the inspiration mechanism. As I will argue more fully in the next chapter, this approach does not take into consideration individual cognitive differences that exist in ideators (Brown et al., 1998). Therefore, the next chapter will present CrowdMuse, a system that can model and adapt to ideators, providing them with more effective inspirations.

CHAPTER 5

CROWDMUSE: SUPPORTING BRAINSTORMING THROUGH MODELING AND ADAPTATION

Online crowds show great potential for creativity. This is in great part due to the large numbers and diversity of participants (Dennis & Williams, 2003; Fischer, 2005). In small groups, the main contributor to an increased performance is synergy, that is, when one person builds on ideas proposed by others (Dennis & Williams, 2003). These synergistic ideas would hardly occur in individual ideation. Therefore, one might expect that by adding hundreds more people to ideation, the likelihood of synergy happening would only increase.

Nonetheless, simply recruiting large numbers of ideators is not enough to ensure a creative output. The same scale and diversity that can boost ideation also presents challenges that hinder the creative output of crowds. The sheer amount of ideas generated can hinder synergistic performance, since an individual is unlikely to be able to read all of the ideas (thus possibly missing the one that could inspire them), much less pay attention to them, which is a requirement for influence (Dugosh et al., 2000; Nijstad & Stroebe, 2006). Therefore, large-scale brainstorming sessions need to be appropriately designed and supported.

Research has attempted to do that in different ways, usually by withholding the entire solution space (all the ideas generated so far) and only exposing ideators to *inspirations*—usually a short text snippet meant to inspire further ideas. These inspirations have taken different forms. For example, Chan, Dang, & Dow (Chan et al., 2016b) employed facilitators to generate inspirations (e.g. questions to promote reflection) during an ideation session. Siangliulue et al. (Siangliulue, Arnold, et al., 2015) attempted to inspire ideators by showing them a small set of ideas that was chosen for its diversity (ideas that differ significantly among themselves) or creativity. Finally, I have attempted to increase the effect of inspirations by adding a small task (e.g. rating the idea) to boost attention to the ideas, as described in the previous chapter.

The common thread between these examples is that they focus on the kind of inspiration being shown rather than on the ideator it is being shown to. In other words: should the same inspiration be presented to two different ideators? Would they be influenced the same? The creativity literature points towards a negative answer to these questions. Theoretical models of idea generation propose that individuals differ on which concepts or categories they generate ideas on (Brown et al., 1998; Nijstad & Stroebe, 2006). This means that each ideator is more likely to focus on some areas (i.e. idea categories) than others. An inspiration strategy that doesn't take that into consideration may be missing out on leveraging their unique strengths for idea generation. For example, if an ideator is more familiar with ideas in category A than those in B, showing ideas in category B may not effectively inspire him or her to come up with new ideas.

In this chapter, I explore how to tailor inspiration selection to individual ideators. The overarching research question explored here is: "*How can we adapt inspirations to ideators in order to improve ideation performance?*" To do so, I present CrowdMuse, a system that models ideators based on their categories of ideation and adapts the system to improve their performance. I begin by describing CrowdMuse, including how it functions in detail, and relating its design back to the literature. I finish by describing two large-scale online studies in which I evaluate CrowdMuse and its adaptive mechanism. In this chapter, I make the following contributions to large-scale creativity:

- I introduce CrowdMuse and the methods it uses to model ideators and adapt to them;
- I validate the system's effectiveness method in two studies, demonstrating that an adaptive system can improve the breadth of idea generation, but this effect is contingent on how the inspiration pool is categorized.

The CrowdMuse System

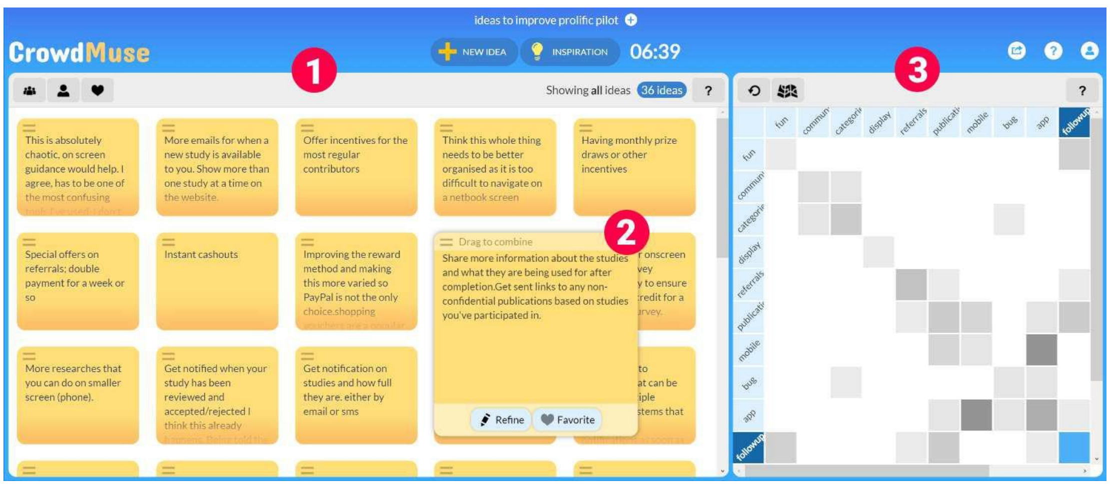

Figure 1. The CrowdMuse system. It has two main views: the idea workspace (1) allows users to view and manipulate ideas by hovering over them (2); and the solution space (3) provides an overview of the density of ideas developed for each tag.

The CrowdMuse system is depicted in Figure 1. It is comprised of two main views. The first, on the left, is the **idea workspace** (#1). The purpose for this view is to allow users to explore and manipulate existing ideas. On the top of the view, a toolbar displays several choices. On its left, there are two buttons, one for displaying all the user's own ideas, the other for displaying the user's favorite ideas. An idea can be favorited by hovering over it and clicking the favorite button (see #2 in Figure 1). On the right of the toolbar, you find a description of what is currently being shown in the workspace (e.g. "Showing *your favorite* ideas" or "Showing *ideas with tag food*", followed by a count of the number of ideas being displayed and a help button (if clicked, a short description of the view is shown).

The workspace enables two other kinds of actions: combining and refining ideas. Ideas can be combined by dragging one idea onto another. This opens a popup showing both ideas, and a space for typing the combined new idea. An idea can also be refined by hovering over it and clicking on the refine button (Figure 1, #2). Doing that, a popup will show up with the idea to be refined, allowing the user to edit its text and submit the updated version. These mechanisms were added in accordance to the principles of brainstorming—in which participants are encouraged to build on one another's ideas (Osborn, 1963)—as well the literature, which has demonstrated the importance of combinations and subsequent iteration (Chan & Schunn, 2015; Dahl & Moreau, 2002; Doboli et al., 2014; Kohn et al., 2011; Ward & Kolomyts, 2010).

The second view is the **solution space**, occupying most of the right side of the interface (Figure 1, #3). By using a matrix form of visualization (Alsallakh et al., 2016), the purpose for the solution space is to provide an overview of which categories have been thoroughly explored and, conversely, those which are yet to be explored. This overview is also important so that ideators are not completely blind to other ideators' performance and can at least try to be more consistent with their tagging of ideas. The solution space is represented as an *n x n* matrix in which the rows and columns correspond to the idea categories developed so far. The color of the cell indicates how many ideas have been developed at the intersection of two categories—the darker the cell, the more ideas have been developed within that intersection. Clicking a cell will open all ideas at that category intersection in the idea workspace.

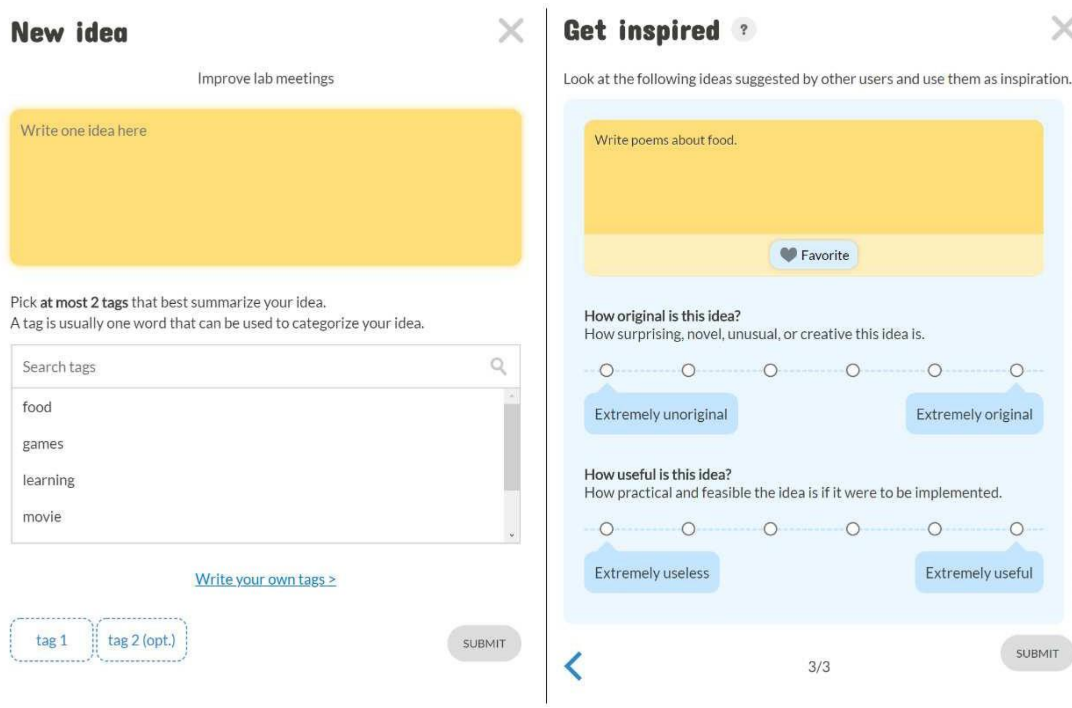

Figure 2. CrowdMuse Popups. The new idea popup is shown on the left and the inspiration popup is shown on the right.

Ideas can be added by clicking the "new idea" button at the top of the UI. When adding a new idea, the user is prompted to pick at most two categories for the idea (based on previously used categories), or to suggest new ones. To the right of the new idea button is an inspiration button, which when clicked presents three ideas along with a small microtask on each (e.g. "rate the idea's originality and usefulness"). The microtasks are used to increase the attention to ideas and consequently their effect on ideators, as described in the previous chapter. This mechanism is designed in accordance with the previous chapter as well as the literature, employing a pull model where inspirations are requested by ideators through the click of a button (Chan et al., 2016b; Girotto et al., 2017; Siangliulue, Chan, et al., 2015). Figure 2 displays both popups.

CrowdMuse as a CST

It is also useful to describe CrowdMuse in terms of the heuristics developed in Chapter 3. The purpose for this analysis is to determine, a priori, some of the strengths and weaknesses of the system as a Creativity Support Tool, similar to what was done for the communication channels in Chapter 3.

Support for Divergent Thinking. CrowdMuse was built precisely for divergent thinking. It supports divergence by allowing ideas to be generated and displayed in parallel in the idea workspace. This contrasts with the usual practice of using a discussion forum, which privileges the latest message in the thread, as well as more in-depth discussions. It also allows ideas to be combined through the drag-and-drop action. Furthermore, it also features the Solution Space view, which shows the different areas that have been explored by all ideators, highlighting empty intersections that are yet to be visited. This intent is corroborated by the evidence discussed in Chapter 5, which showed that the adaptations that CrowdMuse performs can boost breadth of exploration.

Support for Convergent Thinking. In contrast to divergence, CrowdMuse does not provide much support for convergence. The most evident convergent mechanism in it is the favorite button, which can be used as a way of expressing majority preferences for ideas, thus aiding in decision making. However, the system does not currently show a tally of the favorites, therefore reducing the utility of the favorite button for convergence. Another convergent mechanism is the inspiration button, which asks ideators to rate ideas in terms of their usefulness and originality. However, this was added with the only purpose of augmenting the attention to ideas, and therefore the results are not displayed anywhere. In summary, while the system does collect some convergent information, it is not currently displayed in any way, therefore diminishing CrowdMuse's usefulness for convergence.

Support for Shared Material. There is no support for materials other than text in CrowdMuse. You can post URLs (as is often done in discussion forums), but CrowdMuse does not have any features for better supporting them, such as formatting the links or enabling previews of their content.

Support for Shared Understanding. The Solution Space view in CrowdMuse is meant to contribute to the shared understanding of the current ideation space. By showing an overview of density of ideas across the different tags, the system aids in the understanding of where people are focusing, and maybe which areas are yet to be explored.

Support for Collaborative and Iterative Processes. CrowdMuse supports collaboration and iteration through many of the features already discussed above, such as the idea workspace and idea combination. But it more specifically supports iteration through the refinement feature, which allows users to add more information to ideas. This refinement feature was further supported in previous versions of the system through a versioning view, which displayed the history of an idea based on the combinations and refinements that led up to it. The feature was removed due to it not testing well in the usability studies and Prolific pilots.

Support for Group diversity. Since CrowdMuse was designed for short interactions in crowd platforms, a major focus was for ease of use, findability, and learnability, meaning that the system should be able to be easily used from the first moment users interact with it. This can support users with variable technical backgrounds, as opposed to, for example, GitHub. On the other hand, there is not much support for other differences in background.

In summary, CrowdMuse seems to be somewhat the opposite to the channels discussed in Chapter 3. They do not support much divergence, while CrowdMuse focuses on it; They favor converging styles and sharing of different types of media while CrowdMuse does not; they provide some shared material while CrowdMuse does not; they do not have mechanisms to promote shared understanding, while CrowdMuse does; they, just like CrowdMuse, support iteration, collaboration, and group diversity. Overall, however, CrowdMuse's design is mostly focused on enhancing divergence. It intends to do so not only through its design and features, but also through adaptations. This is the focus of the next section, as well as the evaluations done later in this chapter.

Adaptations

The system's purpose is to enhance idea generation by prioritizing categories that could be inspiring to an ideator. This contrasts with current approaches, in which inspiration selection does not take the ideator into consideration, instead being, for example, randomized (such as in the studies described in Chapter 4), chronological (Chan et al., 2016b), or focusing on aspects of the inspiration set (e.g. diversity of the idea set) (Siangliulue, Arnold, et al., 2015; Siangliulue et al., 2016).

As described in the related work, ideators have unique cognitive structures (Brown et al., 1998; Nijstad & Stroebe, 2006). Generally, this means that ideators are more likely to come up with ideas within some categories rather than others, and the ideas may be somewhat temporally clustered together (i.e. ideas of similar categories may be suggested temporarily closer to each other). Therefore, leaving the selection of inspirations to chance may cause them to fail in inspiring (or having as much effect on) ideators due to two factors. Firstly, the ideator is not highly fluent in the chosen categories. This is best visualized with the matrix model. Say, for example, that an ideator has just generated an idea in category A, and from there can switch to categories B, in which she is highly fluent (that is, she has high within-category likelihoods)—and C, in which she is quite inarticulate (that is, she has low within-category likelihoods). In this scenario, an inspiration that touches on category B is much more likely to yield positive results than an inspiration on category C.

The second factor is that the inspiration can break an ideator's train of thought; as proposed by the SIAM model, when ideators generate ideas, they have a concept loaded in their short-term memory (STM). This concept stays loaded until they repeatedly fail to generate more ideas with it. However, if the ideator is exposed to an inspiration that does not match their currently loaded concept, it may interrupt their train of thought, in practice curtailing their fluency within that category (namely, their depth). Therefore, existing research on idea generation shows that inspirations must be carefully chosen to not cause more harm than good.

In practice, the choice of ideas has been shown to influence performance. For example, Siangliulue, Arnold, et al. (2015) compared showing random ideas with an explicitly diverse set of ideas, finding the diverse set to yield greater diversity in idea generation. They also found a set of more creative inspirations yielding more creative results. In the last chapter, I also found that making the set of inspiration ideas similar among themselves yields either no or negative effects. The effect dramatically changed when I changed selection mechanism to be completely randomized, causing the inspirations to improve the breadth of ideation in some cases. Therefore, the importance of choosing the right ideas is not only theoretical, but its effects have already been seen in practice.

Therefore, the CrowdMuse system implements two forms of adaptations: explicit and subtle adaptations. In this subsection, I explain both types of adaptations based on the past literature. Explanation of how these adaptations are powered (e.g. how does the system choose a new category to suggest) is left for the next subsection on user modeling.

Explicit adaptations are designed to be the most influential form of inspiration. They exist in the inspiration mechanism, which presents users with three ideas, each with an accompanying rating microtask. Following previous research, the goal is for the three ideas to be diverse (Siangliulue, Arnold, et al., 2015). However, here the ideas are chosen based on an underlying user model that is generated throughout the ideation session. Each time inspiration is requested, the system will show the user one idea from each of the following categories: 1) an idea of the same category as the user's last generated idea; 2) an idea of a category that is adjacent to the user's current category; and 3) a new category that hasn't yet been visited by the user.

These categories have been selected to curb the two points of failure described in the SIAM model: failure to generate a new idea within the current category, and failure to retrieve a new category (Nijstad & Stroebe, 2006). The first failure is addressed by showing the user their current category, hoping to inspire further ideas within it (effectively increasing fluency within the category). The second failure is addressed firstly by showing users an adjacent category, which is an idea category the user has transitioned to (from their current category) in the past. But it is possible that this adjacent category, which has been visited in the past, won't yield any new ideas. Furthermore, if the system is capable only of suggesting categories the user has visited in the past, it is possible that it would hinder the ideator's breadth by forcing their attention to those categories. Therefore, the system presents ideators with an idea within a category that has not yet been explored by the user. This idea, however, is not random, rather it is based on categories explored by similar users. In other words, the system acts as a recommendation system of sorts, suggesting new categories based on other ideators. This is explained in more detail in the next subsection (User Modeling).

The system also performs an ongoing **subtle adaptation** of the solution space by ordering its rows and columns. This reordering happens every time the user submits a new idea. Since the goal for the solution space is to give users an overview of all developed ideas, the purpose for adapting this view is to guide users' attentions to the most relevant categories when exploring the solution space. The categories are ordered following the same logic as that of explicit adaptations. It orders the solution space, from right to left and top to bottom, in the following way: 1) current category; 2) all adjacent categories, ordered by most to least common; 3) *inferred new categories;* 4) other previously visited non-adjacent categories, sorted by most to least frequent; 5) any other category that has not yet been visited.

In comparison to the explicit interventions, this ongoing adaptation has the advantage of encoding more information, such as allowing users to identify explore overlaps between categories that are meaningful to them. It also retains user agency: rather than pushing three categories deemed useful to the user, they can choose what to explore in more detail. The downside, however, is that with more information being presented at once, ideators are less likely to pay attention to the ideas they encounter and therefore reduce the effect they have on them. This may be particularly meaningful when compared to the explicit inspirations, which employ microtasks to increase attention to ideas.

User Modeling

The adaptations described above are powered by an underlying user model. This model is inferred from a user's behavior within the system. Whenever users add an idea, they are asked to choose one or two categories for their idea (see Figure 2). This selection is done through a list of existing categories, which the system uses to update the user's model. Based on the previous discussion on the adaptations supported by the system, the user model must be able to inform the system about four kinds of categories: 1) What is the user's current category? 2) From the current category, where is the user likely to move to? 3) In which categories is the user most fluent? 4) What are new categories the user has not yet visited but in which they are likely to be fluent?

1) What is the user's current category? This is determined simply by looking at the last idea added by the user. The idea's category is considered to be the currently loaded category. If two categories were used, both are considered to be currently loaded.

2) From the current category, where is the user likely to move? While the user ideates, the system keeps track of category transitions through a *transition graph*: a directed, weighted graph, in which each node represents a category. When the user adds an idea, the system creates an edge between the categories for the latest idea and the preceding ones. The weight of the edge increases as that transition repeats. This is how the system determines the *adjacent categories*.

3) In which categories is the user most fluent? While the user ideates, the system also creates a *category vector* to keep track of the number of ideas the user adds for each category.

4) What are new categories the user has not yet visited but in which they are likely to be fluent? To infer this information, I draw from recommender system techniques (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013). The system uses the user's category vector to identify other ideators that have a similar ideation pattern to their own. The system then calculates the correlation between the user's vector and other ideators'. It selects the top five most similar to make the inferences. Then, for every category the user being analyzed has not yet visited and the other similar user have, the system calculates the average fluency. The category with the highest average is considered to be the one with the most potential.

Study 1

I evaluated the system through an online study on Prolific[8]. In this study, I focus on evaluating whether the combination of the two adaptive features in CrowdMuse improves three well-recognized brainstorming metrics: fluency (number of ideas), breadth (how many categories are surveyed by one ideator), and depth (within category fluency) of ideation. I establish the following hypotheses:

H1. *An adaptive system will increase the number of ideas over a non-adaptive system*. By tailoring inspirations to categories that are more likely to be visited by an ideator, I expect them to be more effective at sparking that new idea that would otherwise have not been generated, either in a new category (thus increasing breadth) or in a previously visited category (thus increasing depth). The result from this is an increased overall number of ideas.

H2. *An adaptive system will increase the breadth of idea generation*. As postulated in H1, an adaptive system will increase overall fluency. I argue that an adaptive system will increase this breadth by showing users their inferred categories, that is, categories they have not yet visited but that are likely to be relevant to them based on similar ideators. In the SIAM model, this equates to delaying the failure in retrieving a new category (Nijstad & Stroebe, 2006).

H3. *An adaptive system will increase the depth of idea generation*. Similarly, I argue that by showing ideas in the current and adjacent categories, an adaptive system will increase the number of ideas they suggest within each category. This will result in an

[8] https://www.prolific.ac/

overall greater depth. In the SIAM model, showing the current or adjacent categories is an attempt to delay failure in the phase of idea generation within a given category or in another previously ideated category (Nijstad & Stroebe, 2006).

Method

I posted a study request through the Prolific platform. Participants were required to have at least 85% approval ratings on the platform, be over 18 years old, and not have participated in any of our previous pilots. All data was collected across three one-day sessions, with two weeks between the first and last sessions. When users accessed the study link, they were shown a short tutorial that went over each part of the system individually. This tutorial also introduced the brainstorming problem, generating ideas to improve Prolific, with the following description: "*Prolific is a great website for researchers and participants alike. However, there is always room for improvement. Come up with as many ideas you can to improve Prolific in any way you can think of. Be as specific as possible in your ideas*". After completing the tutorial, a 15-minute timer would appear on top of the screen and start to count down. After the timer was done, a pop-up screen appeared with a link to a final questionnaire asking about demographics, their experience with the task, and perceptions of the system.

I used a between-subject 2x2 full factorial design in which participants were randomly assigned to a combination of two factors:

1. Solution space (random/adaptive): rows and columns could be ordered randomly, or according to the user's model.

2. Inspiration mechanism (random/adaptive): inspirations could be selected randomly, or according to the user's model.

Idea Pool and Categorization.

Since the adaptive mechanisms are powered by data from other users, I had to pre-populate the system with users and ideas for this first study. This data came from several pilots I ran on Prolific. While the system went through some iterations throughout these pilots (e.g., initial pilots had shorter ideation time), I used the data from pilot participants: 49 users and 189 ideas organized across 54 categories. Categories were determined by the pilot ideators themselves, who could tag their ideas when adding them. I collapsed categories with very few ideas into more popular, broad ones (e.g. the categories *chat*, *forum*, and *email* were all collapsed under *communication*). Thus, when selecting an adaptive inspiration, one of the 189 ideas would be presented based on the match between the desired inspiration category and its own category.

Metrics

I primarily evaluate the effects on brainstorming performance through metrics related to breadth (how many idea categories a user has visited) and depth (the fluency within a given category). To do so, I use two metrics for each of these dimensions. The first metric comes from a manual categorization of the ideas generated performed by me and another graduate student. At the beginning of this categorization, both researchers worked together to define the core categories. As the categorization progressed and no new categories started to appear, the researchers started to work independently (but still co-located), only occasionally discussing where some ideas should be assigned to. 70%

of ideas were categorized in this manner. The remaining ones have been categorized by me based on this initial categorization. Since the majority of the categorization was performed collaboratively, no reliability metric was calculated. This categorization was done blind to the experimental condition in which the ideator was placed. I then extract **breadth** as the number of categories visited. **Depth** is the largest number of ideas a user generated within one category.

Additionally, following the previous chapter, I calculate a metric based on Latent Semantic Analysis (LSA). The vast majority of the corpus for this analysis comes from the same idea pool used in the last chapter, which is comprised of ideas in a similar domain, but related to Amazon's Mechanical Turk (n=7199). I have also added some ideas from the pilots (n=591). This metric uses LSA to build an ideation tree based on the similarity between ideas, in which each node is an idea attached to its most similar parent (see Chapter 4). From this tree, breadth is derived as the number of children nodes of the root, and depth as the maximum number of nodes in one branch. I refer to these metrics as **tree breadth** and **tree depth**. They are calculated just as in the previous chapter.

By using two different metrics, I avoid the bias introduced by a single type of metric. Each metric has its own tradeoffs. I expect the manual metrics to be more accurate, but they are very subjective—different people would likely come to different categorizations. The tree metrics, on the other hand, do not depend on subjective judgements, but they are likely more inaccurate, especially due to the origin of the corpus used to generate them. It should further be noted that both metrics are highly correlated to the user's fluency—someone who comes up with 20 ideas will very likely have higher breadth and depth numbers than someone who comes up with 2, and thus in my analysis I control for fluency.

Results

In total, 115 Prolific users performed this study (42.6% female). Most participants described themselves as non-Hispanic White (75%), with the UK having the largest participation (25%). Participants were randomly assigned to conditions, but since some users would quit the study before finishing it, distribution across conditions was not perfectly balanced. There were 32 participants with neither adaptive mechanisms, 25 with only an adaptive inspiration mechanism, 28 with only an adaptive solution space, and 30 with both adaptive mechanisms.

H1: Fluency did not Change across Factors.

My first hypothesis was that an adaptive system would increase the fluency by increasing both breadth and depth of ideation. To evaluate this hypothesis, I calculated a two-way ANOVA with fluency as outcome variable, and the presence of an adaptive solution space and the presence of an adaptive inspiration mechanism as fixed factors. I include the interaction between factors in the model. There was no significant main effect of the adaptive solution space, $F(1,111)=1.541$, $p=0.217$, or the adaptive inspiration mechanism, $F(1,111)=0.003$, $p=0.957$. I do, however, find a marginally significant interaction effect between factors on the fluency of participants, $F(1,111)=3.74$, $p=0.056$. Figure 3 demonstrates this interaction, along with the mean fluency values.

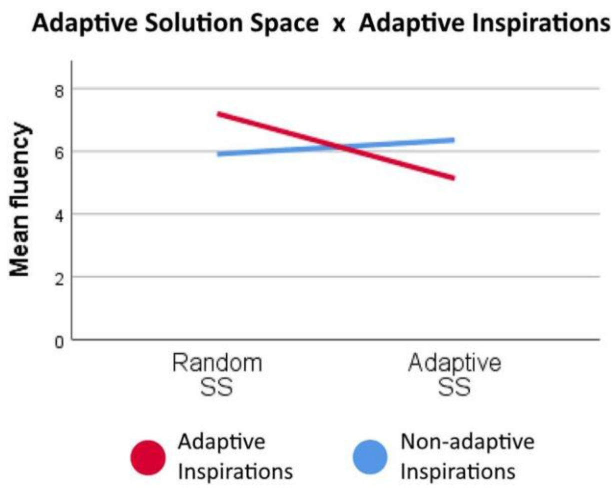

Figure 3. Interaction between adaptive solution space and inspirations on fluency.

H2: Adaptive Inspirations Increased Tree Breadth

I first evaluated the manual breadth metric by running an ANCOVA with breadth as the outcome variable, adaptive solution space and adaptive inspirations as fixed factors, and fluency as a covariate. I included the interaction between factors in the model. I found no effect of adaptive inspirations, $F(1,110)=2.721$, $p=0.102$, adaptive solution space, $F(1,110)=1.482$, $p=0.226$, or the interaction, $F(1,110)=0.358$, $p=0.551$.

I then evaluated tree breadth as an outcome variable. Because there was a significant interaction between fluency and one of the independent variables, I calculated a Mixed Generalized Linear Model (GLM), with the breadth metric as outcome variable, the presence of an adaptive solution space and an adaptive inspiration mechanism as fixed factors, and fluency as a covariate. I included two-way interactions between each factor and fluency. I found a significant interaction between adaptive inspirations and fluency, $F(1,108)=7.949$, $p=0.006$, showing a stronger positive effect of adaptive inspirations on tree breadth as fluency increases. Pairwise comparisons show average and high fluency ideators who were exposed to adaptive inspirations outperformed those with randomized ones. Table 1 details the marginal means for the tree breadth metric across different fluency levels. I found no interactions between an adaptive solution space, F(1,108)=0.706, p=0.402, nor between the two main factors, F(1,108), p=0.286. There were also no main effects of an adaptive solution space, F(1,108)=0.187, p=0.666, or an adaptive inspiration mechanism, F(1,108)=0.308, p=0.308.

Table 1

Marginal means (and standard error) for breadth

	Low Fluency		Average Fluency		High Fluency	
	Random	Adaptive	Random	Adaptive	Random	Adaptive
Breadth	2.55 (0.18)	2.77 (0.23)	4.72 (0.13)	4.98 (0.14)	6.88 (0.18)	7.20 (0.23)
Breadth_{tree}	2.55 (0.24)	2.50 (0.30)	3.85 (0.17)*	4.66 (0.18)*	5.15 (0.24)**	6.82 (0.29)**

* p < 0.05; ** p < 0.00.

H3: No Effects on Depth.

My third hypothesis was that by emphasizing ideas in their current or adjacent categories, ideators would likely be able to generate more ideas within those categories, therefore increasing depth. To evaluate this, I estimated a negative binomial regression, since the data follows a negative binomial distribution. I use depth as an outcome variable, presence of an adaptive inspiration mechanism and presence of adaptive solution space as factors, and fluency as covariate. I found no significant effect on depth of either adaptive solution space, Wald Chi-Square=0.418, p=0.518, adaptive inspirations, Wald Chi-Square=0.211, p=0.646, or an interaction between both, Wald Chi-Square=0.007, p=0.934. The same model with tree depth as outcome variable equally yielded no effect from adaptive solution space, Wald Chi-Square=0.036, p=0.850, adaptive inspirations, Wald Chi-Square=0.294, p=0.588, or the interaction between the two factors, Wald Chi-Square=1.380, p=0.240. Therefore, the adaptations produced no change in depth.

Study 1 Discussion

In summary, I found some support for a positive effect of adaptive inspirations on the breadth of ideation (H2), although only from the tree metric. I did not find significant effects in either fluency (H1) or depth (H3), although I identify a marginal interaction on the former.

I begin this discussion by analyzing the effect of the inspiration mechanism on breadth. Based on the discussion on the differences between the solution space and the inspiration mechanism, it is not surprising that the inspiration mechanism was behind the change in breadth—it was designed to draw greater attention to the ideas (through the rating task), and to only show a reasonable amount of information per request. The question is why it only affected breadth but not depth. One possible explanation is that rating ideas had a metacognitive effect on ideators. Eight ideators have said something to this effect. For example, one user reported that having to rate other ideas encouraged them "*to come up with original and feasible solutions*". Another ideator reported that it "*allowed me to see what users said, and I started noticing patterns*" that allowed her to tell when original ideas came up. This focus on originality of ideas may have pushed users to try to come up with original ideas in different categories.

While I expected the inspiration mechanism to have greater effect, I did not think that the solution space would have no effect. This could perhaps be explained by issues in its usability and comprehension. Many users reported some issue with it such as finding it confusing, finding the tagging poor or redundant, or just having to scroll through so many tags (n=18). In fact, I find that perceptions of how useful the solution space was were correlated with perceptions of the usability of the system ($\rho=0.467$, $p=0.000$) and success in the task ($\rho=0.216$, $p=0.022$), which could mean that those who were confused by it had negative perceptions of the system or the task. Therefore, improving its usability could be key to enabling its effects.

Study 2

Based on the first study results, I made the following changes to the system:

- I recategorized the pool of ideas based on the manual categorization used in the breadth and depth metrics. In practice, this means that the solution space will be better organized, addressing complaints of poor, redundant, or excessive tagging. It should also reduce the noise in inspirations (e.g. miscategorized ideas), improving their effects. We have also included the ideas from Study 1 in the pool of ideas, meaning there are more ideas per category and more user models to draw from. In total, the system now had 899 ideas from 173 users, spread across 19 categories.
- The system now shows the inspirations in the idea workspace after users complete the inspiration tasks. They are colored in blue and with a lightbulb icon to differentiate them from regular ideas. They are also marked when the user clicks a solution space cell. I expect this to increase the effect of the inspirations by allowing users to examine them for longer and more easily refine or combine them.
- I emphasized the refinement and combination actions. I did this in two ways. The first is by focusing on these actions in the intro tutorial, explaining them more clearly. The

second is by adding stats at the top of the idea workspace, mentioning how many original, refined, and combined ideas the user has added. This could help deal with fluency issues seemingly brought on by the adaptive solution space by reinforcing options the user has they see an idea that they were going to add.

Method

In this study, I only examined two conditions: control (inspirations and solution space are randomized) and fully adaptive (inspirations and solution space adapt to users). My main hypotheses remain the same as in the previous study. However, due to the changes described above—especially by having more and better categorized data—I expected a stronger effect in breadth, and possibly some effects in fluency and depth as well.

Study 2 Results

In total, 76 Prolific users participated in this study (40.8% female), 38 subjects in each condition, generating a total of 483 ideas. Most of them described themselves as non-Hispanic White (~61%), with the US (~22%) and the UK (~24%) having the highest number of participants.

I once again evaluate the same hypothesis from Study 1: H1: fluency will increase due to adaptations; H2: breadth will increase due to adaptations; and H3: depth with increase due to adaptations. I use the same statistical analyses employed in Study 1, with the difference being that now there is only one fixed factor, condition. Consequently, I now only evaluate one interaction, between fluency and condition, in the GLMs.

H1: Fluency did not Change Across Conditions

Ideators in the control condition generated, on average, 6.29 ideas (sd=3.07). Those in the adaptive condition generated on average 6.39 ideas (sd=4.29). A One-way ANOVA shows no difference in fluency between conditions, F(1,74)=0.211, p=0.903.

H2: Adaptations Negatively Affected Tree Breadth

An ANCOVA showed no effect of condition on breadth, F(1,73)=1.280, p=0.262. Tree breadth, on the other hand, shows differences. A Mixed GLM shows a significant interaction between condition and fluency, F(1,72)=13.09, p=0.001. However, unlike Study 1, this time the interaction favors the control condition over the adaptive one, showing a marginal difference for average fluency ideators and a significant different for high fluency. Table 2 details how the marginal means change across low, average, and high fluency levels. There was also a main effect of condition, F(1,72)=5.052, p=0.028.

Table 2

Fluency and inspiration metrics for Study 2

Fluency	Control	Adaptive	p
Low	1.88 (0.29)	2.39 (0.25)	0.190
Average	4.48 (0.19)	3.97 (0.19)	0.059
High	7.08 (0.29)	5.56 (0.24)	0.000

Therefore, in this study I find some evidence that an adaptive system hindered breadth compared to a non-adaptive one.

H3: No Effects in Depth

I found no effects of condition on either depth, Wald Chi-Square=0.009, p=0.925, or tree depth, Wald Chi-Square=0.034, p=0.854.

Study 2 Discussion

Like in Study 1, I only found effects on breadth. However, unlike Study 1, this was a negative effect (on tree breadth only) caused by the adaptations. What could explain this difference? I did not find much difference in usage of the refinement and combination mechanics, so it is unlikely that this extra incentive caused such a significant shift. I also find no reason for why the persistence of the inspirations in the workspace could have caused such an inversion of effect.

Therefore, I hypothesize that this striking change in effect was due to the new categorization of ideas. There was a dramatic reduction in the number of categories (54 in the first study to 19 in the second). It is possible, therefore, that the number of categories was too small to be meaningful. Therefore, while in adaptive conditions users were getting a diverse set of ideas, in the adaptive conditions they were being exposed to lesser variety. This could also explain the lack of effect in the manual metrics for both studies. In other words, the categories may be too high level to be significant for both the adaptation and the metrics.

Revisiting the Categorization Scheme

To examine the categorization effect on metrics, I calculated, for both studies, the number of distinct categories users were exposed to through the inspiration mechanism. I then estimated an ANCOVA with the number of categories as outcome variable, condition as fixed factor, and number of inspirations as covariate. For study 1, I find no effect of condition on the number of categories users were exposed to, $F(4,93)=0.535$, $p=0.66$. On average, users across conditions have been exposed to 5.47 ($sd=2.97$)

categories. Despite this, I still found a positive influence of an adaptive system on ideation breadth, indicating that that effect is not due to differences in quantity of categories, but rather on their quality. Study 2, on the other hand, shows a higher number of categories for the control condition (M=4.49, SE=0.19) compared to the fully adaptive one (M=3.85, SE=0.19), $F(2,73)=5.32$, $p=0.024$. A difference in the number of categories users were exposed to could partially explain the change in effect.

The rest of the difference in effect could be due to the chosen ideas being less appropriate for each ideator. If the categories are too high level, and if there are more ideas per category, it is possible that choosing a random idea from within that category may not cause the desired inspiration effects. In fact, I find a marginal difference in how useful ideators in the control (5.82, sd=1.20) and adaptive (5.18, sd=1.66) conditions perceived the inspiration mechanism to be (on a 1-7 scale, 7 being the most useful), $F(1,74)=3.61$, $p=0.061$, indicating a trend towards greater dissatisfaction.

A categorization that is too high-level (i.e. too few categories) could also explain the lack of effects on the manual metrics in both studies, as it may wash off more nuanced category exploration. Therefore, I have also recalculated the manual metrics for both studies. I recoded all ideas for both studies (N=1183) and developed a new categorization scheme with 45 total categories. This new categorization increased the number of categories by breaking down the previous ones. For example, the original scheme had a category called *Study types*, which in the new categorization was broken down into categories such as *collaborative studies* or *in-person studies*. Another researcher then was given this new scheme along with 120 uncategorized ideas and independently categorized them. Agreement between both raters was satisfactory, Cohen's Kappa=0.788.

With this new scheme, I revisited the analysis of the manual metrics in the previous two studies. In Study 1, I recalculated an ANCOVA with the new breadth metric as outcome variable, both adaptations as factors, and fluency as covariate. I now find a main effect of adaptive inspirations on breadth, $F(1,110)=6.200$, $p=0.014$, with adaptive inspiration ideators exploring slightly more categories ($M=6.61$, $SE=0.19$) than those without the adaptive inspirations ($M=5.96$, $SE=0.18$). I still find no adaptive solution space effect, $F(1,110)=0.00$, $p=0.990$, as well as no interaction between the two factors $F(1,110)=0.528$, $p=0.569$. As for depth, I still find no effect of either adaptive inspirations, Wald Chi-Square=0.057, $p=0.812$, adaptive solution space, Wald Chi-Square=0.001, $p=0.976$, or the interaction between the two factors, Wald Chi-Square=0.126, $p=0.722$. Therefore, these results reinforce those obtained through the tree metrics.

I also redid the analysis for Study 2. To evaluate breadth, I recalculated an ANCOVA with the new breadth as outcome, condition as factor, and fluency as covariate. This time, however I find no effect of condition on breadth, $F(1,73)=1.068$, $p=0.305$. I also re-evaluated depth as well, and found no significant effects, Wald Chi-Square=0.034, $p=0.854$.

Discussion

From these two studies, I draw four conclusions: 1) CrowdMuse, particularly through its adaptive inspirations, can positively influence breadth of ideation; 2) The adaptations, as they were proposed, were not capable of improving fluency or depth; 3) The inspiration mechanism had a stronger effect compared to the solution space; and 4) The ways in which ideas are categorized are key to the effect of the adaptations.

The inspiration mechanism's effect on breadth could perhaps be explained by the diversity of ideas presented. In study 2, where the variety of categories was decreased, I found evidence of the system performing as well or worse than control on breadth, potentially because ideas that were too similar were being presented to users. This finding is also in line with previous work, which found that diversity yields diversity (Siangliulue, Arnold, et al., 2015). But, as I also found from a comparison of both studies, a difference in the total quantity of exposure categories does not completely explain this effect, as Study 1 still revealed an advantage to adaptive inspirations despite an equivalent number of exposure categories. Therefore, I argue that the adaptive inspiration mechanism is able to more carefully select inspiration categories and improve breadth of idea generation, but only with an appropriately fine-grained categorization of ideas.

In contrast, both studies showed a lack of significant effects on both fluency or depth. This suggests that the intended effects of the current and adjacent categories were not realized. Their intention was to keep users longer in the current categories, but for both studies I found considerably high likelihoods of users not staying within the same category for two consecutive ideas (95% on Study 1; 85% on Study 2). This lack of effect on depth likely contributed to the overall lack of effects on fluency. It may be that to be effective, the adaptation mechanism needs to better account for the fact that users are likely to switch categories frequently.

I also note that the positive effect found in Study 1 sprung from the adaptive inspirations, not the solution space. As discussed in the system design, I expected that to be the case due to fewer ideas being presented at a time (compared to the solution space), as well as the built-in tasks. Both of these factors should increase the attention to the ideas that were presented, and therefore their effect. However, I also acknowledge that in Study 1, the solution space may have been plagued by usability issues, which may have distracted users from its benefits. I attempted to improve the usability of the solution space for Study 2 by improving the categorization scheme, but also brought its own set of issues.

Finally, a contrast between both studies points to the importance of the categorization of ideas used, both for powering the adaptations as well as for measuring their effects. The two studies showed markedly different results for breadth. I attribute that in great part to the new idea pool used in that study, with significantly fewer categories—although the quantity of pool ideas also greatly increased. As discussed earlier, this may have increased the inner-category diversity, thus diluting the effect of adaptive category selection. Therefore, it is key for an adaptive system to use the right level of abstraction for the idea categories. The same applies for metrics. The initial categorization was not fine-grained enough to capture differences between the factors, which was fixed by the later scheme.

Limitations & Future Work

There are a few limitations with this chapter. Firstly, I note that to build the tree metrics I used a dataset that was comprised of ideas from another domain, as described in

Chapter 4. Nonetheless, their content is rather similar to the ideas generated in this chapter, and it was augmented with ideas from my pilots. Furthermore, I do not rely exclusively on these metrics for my analyses, but also analyze manually derived metrics.

It is important to discuss the quality of the categorization schemes. For Study 1, it came from the crowd itself with minor alterations. The purpose was to approximate the scheme as much as possible to a "real-world" scenario, where there is likely no standard taxonomy of ideas and minimum resources for organizing it. As mentioned, the alterations came from merging smaller categories into larger ones, but only in order to eliminate redundancies (such as those introduced by typos, pluralization, synonyms, etc.). I introduced no new categories in this process. Therefore, that categorization closely approximates one that was organically developed by Prolific workers. As for Study 2, the idea was to generate a categorization from the data. The scheme was developed following an inductive approach by reading the data, developing initial categories, and subsequently consolidating them into broader categories (D. R. Thomas, 2006). By performing this categorization jointly with another researcher, we gain evidence of the clarity of the scheme, as the occasional discussions help ensure the concept for each category is clear. The third and final categorization is very similar to the second, only finer-grained. The approach was to break down larger categories into smaller clusters. The moderately high inter-rater reliability for this third scheme shows that the clarity for the second scheme was maintained, despite the higher number of categories. Therefore, the categorization schemes employed here are appropriate for evaluating CrowdMuses's adaptation features, as they either emerged from the community itself or were performed according to common qualitative coding practices.

The contrast between the two studies presented here provide some information on the importance of the categorization for the adaptations to work. Namely, I found that a smaller pool of ideas spread between a larger number of user-generated categories was able to positively affect breadth. When increasing the number of categories and decreasing to fewer, researcher-generated categories, the effect was inverted. This is a useful finding for the design of such adaptive systems, but there is still more to be understood before reaching conclusions on best practices for the categorization scheme. Since I only compared two different schemes, there is not enough information to infer how the effects progress across a range of category numbers. Therefore, more work is needed to elucidate how the effects change as the number of categories change, and whether that is generalizable to other domains and user pools. Furthermore, there are also limited inferences I can make on the nature of categories themselves, rather than simply their numbers. In study 1, categories were user generated (with minor adjustments), while study 2 categories were generated by the researchers. The impact that this difference may have caused across both studies is unclear, and more work would be necessary to understand it. These factors are key for CrowdMuse's usefulness in a real-world context, in which the categorization scheme would frequently change, especially at the earlier phases of idea generation, and therefore should be systematically evaluated in future work.

Finally, future work could probe further into the synergistic effects between both types of adaptations. I find some evidence of such effects in the marginal differences in fluency in Study 1. I also argue that the effect of adaptations could be further enhanced by using different types of inspirations. In the related work, I described research that explored ways of inspiring ideators more efficiently than just idea exposure, such as stimulating questions. Future work could apply those techniques in conjunction with an adaptive selection mechanism in order to improve their effects. Furthermore, research could explore how to adapt to users not just in the categories of inspirations, but also in their types, by, for example, adaptively choosing whether to inspire ideators through an example, a schema, or posing a thought-provoking question.

Conclusion

In this chapter, I presented CrowdMuse, a novel system that models and adapts to users to improve their ideation performance. I evaluated the system in two online studies. I found that the adaptive system was capable of improving breadth of ideation, specifically through its inspiration mechanism. The adaptive solution space did not affect results, though issues of usability may have affected its effectiveness. Neither depth nor fluency were affected by adaptations. Finally, I also discussed the effect that categorization schemes of varying levels can have on the adaptations as well as measurements.

However, the study presented in this chapter, just as many other studies in the crowdsourced creativity literature, was performed in an artificial crowd rather than in communities who could benefit from these findings. Therefore, in the next chapter, I describe the deployment of CrowdMuse in an existing online community, thus connecting this work back to the one presented in chapter 3. A qualitative analysis of this deployment can point to how crowd creativity systems described in the literature could be deployed to elicit the creative power of existing online communities.

CHAPTER 6

DEPLOYING CROWDMUSE IN AN ONLINE COMMUNITY

In Chapter 3, I examined the communication channels currently used by some online communities for their creative collaborations—discussion boards, mailing lists, and bug trackers. Through that analysis, I uncovered different bottlenecks that the affordances of such channels could bring to creative collaboration. Chapters 4 and 5, on the other hand, focused on advancing the research on crowd-scale creativity by proposing ways of improving impact through enhanced attention and appropriateness of inspirations. One of the outcomes from that research is CrowdMuse, a system that could help facilitate distributed ideation in communities such as the ones examined in Chapter 3. Since it was designed with brainstorming in mind, it could possibly help communities overcome some of their creative bottlenecks. Other researchers have created similar systems, such as the previously discussed IdeaHound (Siangliulue et al., 2016) and IdeaGens (Chan et al., 2016b), which could similarly have the potential to help distributed groups overcome creative bottlenecks in collaboration.

Nonetheless, these systems are usually disconnected from the practical contexts in which they could be useful. Usually, these systems have been designed for and evaluated in large-scale crowd markets such as Mechanical Turk. This brings a few implications to their design. First, it means that they will be used in a single, non-repeated, short interaction. Usually anywhere between 7 and 30 minutes. This also indicates that many people will perform the ideation task motivated by the promised payment rather than due to intrinsic motivation. Finally, due to the diverse (and many times unknown) backgrounds of crowd workers, the problems have to be very general so that most people can participate. Aside from these issues, these systems also usually attempt only to improve idea generation, but not the sequential steps such as refinement and selection. Therefore, the crowdsourced creativity endeavors discussed throughout this entire dissertation may face obvious barriers if they were to be adopted by online communities as a solution for their creative needs. In other words, their designs may be appropriate for Mechanical Turk or Prolific, but not for established online communities.

I do note, however, that the goal for some of these approaches is, in fact, to use crowd markets as their source for creativity, or to promote specific workflows rather than adapt to existing ones. For example, the BlueSky system uses a "mechanized creativity" process, in which a very specific workflow is distributed to crowd workers in order to exhaustively explore the solution space (Huang & Quinn, 2017). Approaches such as this could be considered process-centric approach, in which a system enforces a specific creative methodology onto a given group of ideators. This means that it may sound somewhat unfair to judge such approaches on the basis of their fitness to online communities. However, I take the position that large-scale creativity finds its greater strength by augmenting current practices. At the core of this position is the fact that motivation and knowledge lie at the core of creativity (T. Amabile, 1983). Also, "forcing" a process onto individuals or communities may fail to account for differences in individuals, such as those described by theoretical models (Nijstad et al., 2010, 2002; Paulus & Brown, 2003) and not trigger the proper motivational cues (T. M. Amabile, 1985). But perhaps most importantly, individual creativity is a key component in broader level creativity models (such as creativity in organization; T. Amabile, 1988). This is unlikely to be the case in situations where individuals are simply assigned tasks as part of an automated process. Therefore, this individual or community-centric approach can only be accomplished by understanding and adapting to current practices.

This exploratory chapter begins to look into this issue, particularly as it relates to CrowdMuse's design and features. By deploying CrowdMuse in a FOSS community for two weeks, I explore the following questions:

1. How do online communities use CrowdMuse?
2. What community needs does CrowdMuse address?
3. What community needs does CrowdMuse not address?

By analyzing the three points above, we may start to understand the strengths and weaknesses of current large-scale creativity approaches. Therefore, this chapter contributes to the research on support tools for large-scale creativity in the following ways:

1. It describes differences in the usage of a brainstorming system between crowd markets and established online communities;
2. It details needs that emerged from the deployment, which could guide the design of large-scale crowd creativity tools that are designed to support online communities.

While the work done in this chapter is highly exploratory, it can be informative for designers and researchers of crowd creativity systems, as it indicates that much of the focus of creativity research may not be aligned with the perceived needs of online communities.

Deployment Community: FreeCAD

For the target of this deployment, I chose to revisit one of the communities examined in Chapter 3: the FreeCAD project. This choice was due to the level of activity within the community forums (as described in the next paragraph), the management culture (very open as opposed to something managed by a foundation or enterprise), as well as my previous experience with it—it was one of the only communities from which I was able to get interview participants from in Chapter 3.

FreeCAD[9] is an open-source CAD tool for designing 3D models of real-life objects. The project started in 2002 and has seen continuous support since then. At the time of this writing (Nov. 2018), its currently release was 0.17, and its GitHub page[10] indicated 15,181 commits (which only started in 2011, when they migrated source control from SourceForge to GitHub) from 175 different contributors. Participation in their discussion forums is high as well. Again, at the time of this writing, its discussion forum indicated a total number of 18,277 registered members, with a total of 254,464 posts across 22,761 topics. Therefore, the project is quite active in both its development and discussions.

FreeCAD's Creative Processes

Chapter 3 detailed how FreeCAD and other FOSS projects propose and discuss new ideas through communication channels such as discussion forums and what impact these channels could have in the creativity of the outcome. Usually the process starts with

[9] https://www.freecadweb.org
[10] https://github.com/FreeCAD/FreeCAD, accessed on 11/09/2018

a well-defined idea proposed by a user, which tends to lead the thread to focus on whether that one idea is good or not—a clear limitation on divergence.

However, there are some instances members specifically open a topic to brainstorm new ideas for something. During this deployment, I was directed by a community admin to a topic in which a member proposed a brainstorming around ideas for features and improvements to be included in FreeCAD's next milestone. This was the initiative of one member, who described the brainstorming process and enforced it throughout the brainstorming period. Given that this thread provides us with an organic example of community members running a strict brainstorming session through the discussion forum, it can provide insights on the shortcoming of current methods as well as the challenges involved in doing so. Therefore, I will refer to this thread in the results section as the *Milestone Brainstorming*.

Method

Before describing the deployment, it is important to note that the system underwent changes compared to the studies described in the previous chapter. These changes were made in order to better support a more natural context. First, I added a "Load all ideas" button to the idea workspace toolbar, allowing users now to see all the ideas at once. Second, every new idea added will be added to the pool, meaning that users can see each other's ideas. Third, all users are now placed in the adaptive condition.

Deployment began by contacting the one of the administrators of the project. This was done both to ask for permission to post in the community, as well as to query about a topic that might matter to the community. The admin responded positively, and suggested three topics:

1. How to help accelerating the integration of Assembly3 into FreeCAD
2. How to help the adoption of FreeCAD in schools and universities
3. How to attract new developers to FreeCAD

I chose the second topic for two reasons: 1) its less technical nature (compared to the first topic) may allow more people of diverse backgrounds and experience level to contribute to this brainstorm; 2) Since CrowdMuse does not support features related to code (such as syntax highlighting) or other types of media, avoiding discussions around coding may be more appropriate.

I then made a post in the “Open Discussion” sub-forum announcing the study. This initial post contained: 1) a description of the brainstorming problem; 2) a link to the system; 3) the deadline, set for one week after the post was made; 4) rules for the study, which are that all ideation-related tasks must take place in the system, and that the thread can be used to discuss ideas or the study; 5) compensation, taking the form of a donation to the project. This donation is proportional to the number of participants, $5 each, up to a limit of $250; 6) Privacy information. During this first week, I made another change to the system, which was to allow numerical characters in the tags. This was a request made by one of the community members.

After the first week, I posted the final questionnaire. However, as suggested by some of the community members, as well as due to a low engagement with the system and the discussion thread, I extended the study for another week. As part of the extension, I also proposed some discussion questions for the community in the thread, as well as a $50 amazon gift card for those who participated in the discussion.

After the study was concluded, I contacted those who filled out the final questionnaire and invited them for an interview. I also contacted the author of the milestone brainstorming thread, since he had experience conducting a formal brainstorming session within the community and therefore could have valuable contributions for my research goals. Two members declined the interview, and the other two, as well as the author of the milestone thread agreed for the interview.

In summary, the data available for my analysis was: CrowdMuse system logs, messages in the study and milestone threads within the FreeCAD discussion forum, feedback form responses, as well as 3 interview transcripts. The logs were parsed in order to obtain a quantitative summary of the behavior of participants in the system. To analyze the remainder of the data, I followed a general inductive approach as outlined by D. R. Thomas (2006). The data—including feedback forms, discussion forum posts, and interview transcripts—were put into a spreadsheet. Following a reading of the data, I began the creation of the first categories, which yielded a total of 45 categories. I then parsed through the categories, joining them into the main topics that will be discussed below.

Results

Demographic information

Participants who filled out the final questionnaire (P1, P2, P3, and P4) were all male, with ages 25, 54, 30, and 58, respectively. Two of them were from the US (P2 and

P3), one from Spain (P1), and another from Germany (P4). Three defined themselves as non-Hispanic whites (P2-P4), and one as Hispanic or Latino (P1). Their backgrounds were in Industrial (P1), electrical (P2), and biosystems (P3) engineering, and one in computer science (P4). They have participated in the community for 5 years (P3), 4 years (P2 and P4), and 1 year (P1). The most junior of these contributors (P1) estimates spending around 2 hours a week contributing to the project, while the other estimate between 10-20 hours a week. Two of these (P3 and P4) are moderators in the community. From the users who filled out the questionnaire, I have interviewed P1 and P3. In addition to them, I have also interviewed P6, who created the milestone thread but did not participate use CrowdMuse (I instead showed it during the interview). P6 described his background as having worked in corporate IT for 30 years (and now retired) and having contributed to the FreeCAD for about 4 years. He is also male, US based, non-Hispanic white.

In summary, this sample is entirely male, mostly with an engineering background, mostly white, and mostly having participated in the project for a considerable time.

CrowdMuse Usage Summary

By the end of the second week, 11 different users used the system, generating a total of 21 ideas (an average of 1.9 ideas per user, sd=0.89). The majority of ideas (16) were added during the first week. The average idea length was of 318 characters (sd=264.23). Of these ideas, 18 were original and 3 refined. These refinements were made for different reasons. One was meant to fix a typo. The other two were meant to add more information to the idea (by simply adding a new sentence rather than changing the original text). There were no ideas stemming from combinations. On average, each user clicked on solution space cells 12.82 times (sd=10.26) and completed inspiration tasks 1.18 times (sd=1.19). Over the two-week period, users visited the website an average of 2.1 times (sd=1.44, max=6, min=1). The favorite button was not used much, with 3 different users favoriting 4 ideas altogether.

Table 1

Usage statistics across Chapter 5 and 6 Studies

Metric	C5: Study 1	C5: Study 2	Deployment
Users	115	76	11
Fluency	6.10 (3.51)	6.34 (3.7)	1.9 (0.89)
Elaboration	94.58 (69.41)	99.28 (91.46)	318.00 (264.23)
Refinements	0.39 (1.01)	0.91 (1.38)	0.27 (0.44)
Combinations	0.30 (0.75)	0.96 (1.46)	0
S. Space clicks	14.78 (13.65)	19.14 (12.09)	12.82 (10.26)
Inspirations	2.40 (2.55)	1.84 (1.75)	1.18 (1.19)
Visits	-	-	2.1 (1.44)

To contextualize these numbers, Table 1 summarizes some of the main usage metrics across the two studies reported in the previous chapter, placing them alongside to the ideation session described in this chapter. On average, usage metrics seem somewhat lower than those of the deployment, but the differences are not so marked in some of them: inspiration requests, solution space clicks, and refinements. No one used the combination feature, although that could be due to the low sample size, since that feature was not very used in the crowd studies either. However, the most striking difference appears in fluency and elaboration (characterized by the number of letters in an idea). While FreeCAD users had much lower fluency (~2 vs. ~6 in Prolific), the difference in an idea elaboration was quite marked (318 vs ~100 for Prolific). A closer examination of the ideas shows them to be more grounded and detailed than those generated in the crowd market. While these differences may point to a trend, it is important to note that given the very small *n* as well as the numerical differences in the pool of ideas (a couple dozens of ideas vs. hundreds of them) and methodological differences (there was no mandatory time duration for ideation), these differences could change with larger participation.

The 16 tags generated by the community were: modelling, inperation [sic], books, course plan, workbench, workshops, seminars, contest, promotion, web, startup, research, early education, UI, admin, and security. Figure 1 shows the final configuration of the solution space as of the end of the second week. The lack of a well-defined diagonal indicates user preference for multiple tags rather than single ones. In fact, users often suggested the desire to tag ideas with more than 2 tags. The darker cells represent a density of 2 ideas, while lighter ones represent a density of 1. Therefore, the range between the most and least popular tags is minimal.

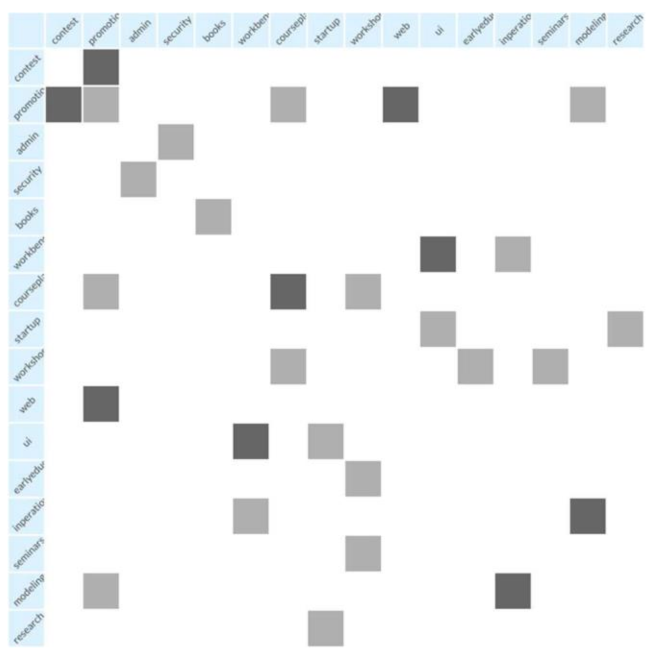

Figure 1. Final configuration of the solution space. Darker cells represent a density of 2 ideas, while lighter ones represent a density of 1.

As for the discussion, 8 different users posted in the forum thread. During the first week, there were a total of 10 posts, of which 7 were asking about system issues or features (e.g. being able to use numbers in the tags, fixing typos) and 3 were talking about the task itself (e.g. suggesting longer duration for the study). During the second week, there were 7 posts, of which 5 were discussing the questions I proposed at the beginning of the second week, one was discussing the system, and one was asking whether the study was still open.

I now turn to the user's perceptions of CrowdMuse in terms of its perceived benefits as well as the issues it fails to address.

Perceived Benefits

Users expressed having **benefitted from the solution space**. P6 reflected on his experience of having to categorize the ideas during the milestone thread: "*Yeah, it was difficult. Basically, it was just looking at the list and saying "that sort of goes with that, that goes with that". Got them into clumps and tried to come up with a descriptive title to the clump.*" Due to that experience, he appreciated having the solution space: "*Having the crosstab on the other side happen automatically rather than having to figure out how to put it on a spreadsheet is great*" (P6). P3 also praised that feature: "*You had the chart with two axes, I thought that was really interesting way to see the relationship between issues*" (P3), though noted that being able to see ideas without having to click on the cells would be better. P4 also enjoyed it and wished that he could "*drag elements from the* *solution space to the left in order to use them for refinement*" (P4). P2 also responded "*dark squares*" when asked his opinion on the most helpful part of the system.

The appreciation for the solution space likely stems from a **need to develop shared understanding** of all proposed solutions, something which is not supported too well by the discussion forum. P1 says he liked the solution space because "*it allows me to have an idea about what are the main points that I should focus on [...] So in that view I can look at a tag and say 'Ok, people think that those are important'*". He contrasted that to the discussion forum: "*I think it's really easy to forget things, older posts, and you miss a lot of things if you don't keep up to date and spend a lot of time with that*". This kind of summarization had to be done manually in the milestone thread, as reported by P6: "*I periodically summarized them [the ideas] into a spreadsheet just so people could get their head around what had been talked about in the thread*".

Users also appreciated CrowdMuse's **structure for displaying ideas**. P1 contrasts it with the discussion forum: "*In my opinion CrowdMuse is a better way to brainstorm than a linear tool such as the forums*". This relates back to the discussion of creativity bottlenecks in Chapter 2, which identified the discussion forum's linear structure as a potential bottleneck to divergence. In fact, P1 highlighted CrowdMuse's potential for early phases of idea generation: "*for me, it's quite useful, because you can take a lot of input, just mix it. So I would mainly use for the beginning, when I have an idea, a project I'd like to start, in order to get some initial feedback, I think it's perfect*". P3 contrasted it with the discussion forum: "*It's specifically designed for that sort of thing, whereas the forum is designed for general discussion. And that sort of iterating on the concept of improvement, features, and so forth is something that happens within the* *forum along with other things but having something specifically for that would definitely be useful*".

Designing participation around ideas rather than a discussion can also **curb bad behaviors** commonly seen in the community. A strong complaint made by P3 was bikeshedding, which is when a thread is hijacked by discussions around unimportant matters, which draw away from the main purpose of the thread. P3 says: "*Our current method for hashing out ideas can be derailed by bikeshedding, but this looks like it could help avoid that.*" He says that "*the first two or 3 pages of a thread about a particular issue are pretty constructive […] the thread can go on for 8 pages or something like that where maybe 2 or 3 people arguing back on forth on less important details, rather than the overall idea itself. That doesn't seem that fruitful*". P6 also complained about an overall inclination for people to just argue over anything: "*Be prepared for people to object to everything. Or anything you come up with.*". P1 also argued that CrowdMuse is perfect for some tasks as there is no "*censoring*". Therefore, the way CrowdMuse structures collaboration is beneficial to reducing some unwanted behaviors, although at its current form it also brings drawbacks, as will be discussed later.

Users also enjoyed the **idea manipulation features**. P3 summarized it well: "*something like CrowdMuse could be nice, 'cause someone could go in there, in sort of a organic way, where they can suggest their ideas, someone can combine it with another previously submitted issue or suggestion, and then can be refined from there, and people can see what are the popular feature suggestions*". P6 expressed liking the combination feature. They did, however, suggest improvements. P1 would like to be able to review the refinements made on his ideas. He also suggested being able to divide ideas into two, as sometimes they may get too large. I also note that the manipulation features in the system were very seldomly used. This means that while conceptually users may enjoy having such features available to them, it doesn't mean that will use them.

Even though the **authentication system** used in CrowdMuse had its design simply due to simplicity for running large-scale studies, it prompted some unexpected feedback from users. In fact, the first response I got in my study thread was P4 asking how users are identified in the system. Simply put, when you access CrowdMuse's URL for the first time, it generates a random ID and keeps track of it via a cookie for later visits. No sign up or identification is necessary. P3 was particularly excited about the low friction this system enabled: "*in order to become a contributor or participant, you have to go in there, create an account, create a post that needs to be approved by someone. There's some friction there. So having a system that has very low friction in the way to contribute can be very helpful*". This problem of low participation from newcomers was also approached by P1: "*The problem I think I see is that people that already have been using FreeCAD for a while, they know how it works, so they don't see as many problems as newcomers see, and newcomers do not participate that actively*". P1 further argues that this would be less of an issue if "*feedback was easier to present*". Therefore, very low entry barriers such as CrowdMuse's authentication system could benefit the diversity of participation within the community. Naturally, there are drawbacks to this approach, such as multiplatform users being unable to use the system under the same account in different platforms, which would make a user login feature necessary (P1).

Important or Unaddressed Issues

Despite the benefits users perceived CrowdMuse to bring, there are also areas in which CrowdMuse does not provide enough or any support, despite the community's needs. I turn to those issues now.

Perhaps the most salient issue is the **lack of convergence mechanisms**. This is most evident due to a lack of a discussion feature, which is at the core of their current practices. In fact, P3 described their process: "*We try to aim for rough consensus, but our process proceeds very organically*". This is similar to P2's account: "*For us it's usually long forum threads, until we either build consensus or someone "takes the bull by the horns*" . Therefore, it would make sense that some discussion would be missed by users. P1 said: "*I may miss [in CrowdMuse] a little bit of feedback over the ideas that I input*" P2 goes further: "*I am a big fan of the brain storming process. Perhaps our inexperience refining the ideas influences my impression that CrowdMuse didn't seem to build consensus?*". Therefore, the lack of discussion in CrowdMuse was seen as an issue for community adoption. Aside from discussion, users also suggested a voting mechanism ("*It's useful to be able to vote for other ideas, it's another way of input, not just commenting, I like it I don't like it*", P1) as well as some general usefulness metric ("*Have some sort of data on how many would find it useful.*", P3).

While CrowdMuse's solution space and tagging system was welcome, there were still **issues with the tagging mechanism**. The issue of tagging or categorization is both important and difficult. It was at the core of P6's reasoning for creating the milestone thread: "*It was pretty disorganized forum opinions, and arguments, there's no structure here at all. We need a little bit of structure in the discussion*". His approach then was to create the thread and categorize the ideas proposed in there. When asked about the categorization in the milestone thread, P6 said: "*Yeah, it [the categorization] was difficult*". As for how well CrowdMuse supported this, P1 mentioned several times that he wished the system allowed him to add more than two tags to an idea: "*I'd like to enter more tags. […] That would be a good improvement, because it would give me more flexibility*". Furthermore, CrowdMuses' categorization is not automatic, and therefore may still require some management. It currently does not support much manipulation of the tags generated by users, and therefore a more advanced categorization system could be necessary for it to meet the community's needs.

Another factor that is not appropriately supported in CrowdMuse is **leadership and decision making**. As described by P2, their discussion goes on "*until we either build consensus or someone takes the bull by the horns*" (P2). When neither of these happen, it simply brings many discussions to a halt—no decisions are ever made, and new features are not developed. P3 gives an example: "*For example, we have an issue where a feature I wanna implement, and other people see the value, but people who are more experienced in the community don't really see the value for it. So we've reached a stalemate in the community*". He also mentioned another core feature that been in the oven since around 2011 because they couldn't decide how to move on. He compares this to the Python community, which has "*benevolent dictator for life, who can make the decisions*". P6 agrees with him: "*But yeah, you need somebody to play that role 'sorry I've listened to all of you so that's how it's gonna be'. Dictator is not always that bad*".

The challenge, even after consensus is reached, is to **transform the ideas into actionable items**. In this sense, P4 reflected about the milestone thread: "*I think it was* *our best attempt yet [at a brainstorming session]. The challenge is always to turn data into information*". And P6, who originated and managed the thread, agrees with that assessment. When asked about what was the hardest part in managing that thread, he said: "*The transition from "here are all the things we talked about" to actually developing work packages, assignments*". P3 adds more detail: "*For example, once a feature is decided and has been kinda refined, it would be nice to have […] something very concrete, something ready for a developer to work on, something that is actually actionable rather than a nebulous idea*". Therefore, it is important for these community members that a brainstorming actually yields concrete products that can be handed over to a developer.

A final issue that relates to many of the points above is that of **limited resources** for admins to manage a dedicated brainstorming tool. P3 says: "*the downside [to a dedicated brainstorming tool like CrowdMuse] is that it may require extra administrative effort and attention*". Asked whether this extra effort could be worth it, he replied: "*I would say so. The deciding factor which could flip my answer the other way is that our sysadmin resources are limited--I would be the one to create and maintain the resource, so if I got around to doing so and, for example, saw something that suggested it would be a headache to maintain, or might have security problems, I might change my mind*" (P3). CrowdMuse does not require much maintenance, but in some part due to some of the issues discussed above. For example, it has no convergence mechanism, which would surely require some overseeing. Therefore, if such a feature were implemented, it should be designed to minimize the need for management.

Discussion

A comparison of the behavior of users in this deployment shows that while most metrics showed lower numbers in the deployment compared to the crowd market used in the previous chapter, the most marked differences are in fluency (~6 in Prolific vs. ~2 in FreeCAD) and elaboration (~100 in Prolific vs. ~320 in FreeCAD). Therefore, while FreeCAD members posted considerably fewer ideas on average, their ideas were much more well developed than those of Prolific users. This could stem from FreeCAD user's deeper knowledge of the project and its needs, especially given the admin representation in my sample. Another explanation is that members of the FreeCAD community like to write a lot, as pointed out by P3: "*It's the people who like to talk [who participate in the forum]. If you give them a topic to talk on, […] they'll take that opportunity to talk for as much as they feel like*". Most crowd brainstorming systems so far tend to focus on (or produce) short text snippets. Therefore, one take away from this deployment is for system designs to expect longer and better articulated ideas from communities such as FreeCAD.

On the other hand, there is also the issue of low fluency. Users generated close to 2 ideas each. While we have a small sample, this pattern was not much different in the milestone thread, in which 30 users generated an average of 1.76 idea posts each (although sometimes they would conflate more than one idea in the same post, which could bring this average somewhat up). However, participants expressed belief that the community was creative in general. When asked what the easiest part in the milestone thread, P6 was emphatic: "*The easiest part, 'cause I didn't have to do it, was coming up with the ideas. There were lots of people who had ideas and wanted to contribute them*". In general, participants seem to agree with the notion that the community is creative and full of ideas. Therefore, they do not seem to perceive creativity or idea generation as an issue, despite showing really small fluency.

Despite not perceiving idea generation as an issue, it is important to distinguish between perceived and actual performance. And the numbers indicate that users tend to have low fluency (and, consequently, breadth and depth) in this community. Therefore, communities such as FreeCAD could still benefit from features aimed at increasing fluency and divergence. However, the mechanisms usually employed by large-scale creativity systems may not be the appropriate choice for that, at least not by themselves. Systems like CrowdMuse address the issue of accessing new categories, but they tend to assume that the ideator is searching for as many categories as possible. FreeCAD users, on the other hand, seemed to optimize for very few but very detailed ideas. Therefore, a possible way of improving their diverging performance would be to explicitly structure the ideation task into divergent and convergent phases. In the divergent phase, users would be queried to come up with short snippets for their ideas. In the subsequent convergent phase, they could add more details to their ideas.

As for the themes that emerged from the interviews, participants perceived value in CrowdMuse due to: the solution space, their need to develop shared understanding, its structure for displaying ideas, its potential for curbing bad behaviors, its idea manipulation features, and the low friction authentication system. But they also highlighted issues that are important and are not necessarily covered by CrowdMuse: its lack of convergence mechanisms, its limitations with the tagging mechanism, no support for leadership and decision making, no way of transforming ideas into actionable items, and concerns with the limited availability of resources admins in the community have.

Table 2

Perceived benefits and issues and their relation to CST heuristics

Metric	Benefits	Issues
Divergent Thinking	Structure for displaying ideas; Idea manipulation features;	-
Convergent Thinking	-	Lack of convergence mechanisms; No support for leadership and decision making;
Shared Material	-	-
Shared Understanding	Solution space; Need to develop shared understanding;	-
Collaborative and Iterative Processes	Idea manipulation features;	Limitations with tagging; No support for leadership and decision making; No way of transforming ideas into actionable items;
Group Diversity	Curbing bad behaviors; Low friction authentication;	Limited admin resources;

To better understand the core benefits and issues CrowdMuse was perceived to have, I categorize them according to the list of CST heuristics developed in Chapter 3 (Table 2). I add some of the items in more than one category, as they may concern both heuristics.

This shows that CrowdMuse's greatest strengths lie on supporting divergent thinking and shared understanding, while its greatest limitations lie in the domain of convergent thinking and collaboration. These results would likely be the same if any other tool such as IdeaHound (Siangliulue et al., 2016) or IdeaGens (Chan et al., 2016b) were the subject of this deployment, as systems in this domain usually focus on idea generation. This is an broader issue in the creativity domain, as CSTs have historically focused on the ideation phase in lieu of other phases such as problem finding or idea evaluation (Gabriel et al., 2016; K. Wang & Nickerson, 2017). Therefore, these results add further evidence to the fact that creativity support systems need to start exploring how to integrate ideation support with other phases of the creative process. This is especially true for systems that are to be used in communities with a tradition of discussion (to the point of excess) such as FreeCAD.

On a final note, none of the comments related to the shared material heuristic. However, this does not mean that that is not an important form of support. Looking at the milestone thread, we find that 25% of threads used some form of external material, including images, videos, CAD files, spreadsheets, PDFs, and URLs. That fact that nothing related to shared materials was mentioned may be due to CrowdMuse not supporting discussion, thus reducing the need to expand on points using some form of shared material. Therefore, supporting shared materials would likely be important for a creativity support system to be useful for a community such as FreeCAD.

In summary, this deployment points to the need for large-scale creativity support systems to: 1) expect users to come up with larger, more detailed ideas; 2) possibly explicitly support divergence by, for example, diving it into two phases; 3) explore how to support other phases of the creative process in conjunction with idea generation.

All of this prompts a discussion on the adoption of new practices by the community. If systems such as CrowdMuse propose support for creative processes that differ from the community's own processes, how we can enable adoptions of these new practices? How can we carry the benefits found in crowd market studies to real world communities? The milestone thread may provide some information in this regard.

Looking at the thread, you find that it was successful in P6's intent of following a strict brainstorming process, despite the community's lack of practice with it. But this was not by chance. There are visible attempts to divert from this process, which were curbed by P6. He expressed that this was the most difficult part of the process. Therefore, there may not be immediate user buy-in of a new process, but the presence of a known community member in enforcing these rules may be a key component of compliance. However, as expressed by users, time is limited, and management of brainstorming processes is not easy. Therefore, adoption could be reinforced by a perception that such systems are saving their time in the process of brainstorming, refining, and developing ideas into workable items that can be readily tackled by a developer. Finally, user agency needs to be preserved. As noted, often ideas in open-source communities spring from a user's individual experience with the system. This results in generally well-developed ideas from the start, which clashes with the notion of breadth before depth championed by the brainstorming technique. A compromise could be to structure idea input into different parts, primarily between the "gist" of the idea and its details. The system would allow users to input both but hide the details from others until other users have developed their own ideas. This way, the system would allow users to follow their instincts, while still fostering greater divergence. Future work could probe into these points to better understand ways of facilitating community adoption of these new tools.

Limitations

It is important to note the limitations with the research presented in this chapter. First of all, this deployment was made in only one community, with a small, relatively homogeneous and highly-engaged sample. This prevents any evaluation of CrowdMuse's adaptive features. This also presents clear implications to the generalizability of these results to other communities, or perhaps even to more peripheral members within the FreeCAD community. Other communities, with their idiosyncratic processes and culture, could and likely would respond differently than the FreeCAD community did. Many of the communities surveyed in Chapter 3 have a more centralized or organized management style. For example, LibreOffice is managed by the non-profit The Document Foundation, while Visual Studio Code is managed by Microsoft. Therefore, topics such as leadership may play a more prominent role, while concerns with resources may be reduced to an extent. Nonetheless, I do believe that these findings are relevant to other communities as well, especially in light of the bottlenecks discussed in Chapter 3. Other communities, especially those who share some similarity with FreeCAD, struggle with the same issues, and therefore it is likely that they could perceive the same benefits from CrowdMuse as our sample of the FreeCAD community did.

The fact that CrowdMuse was the only system deployed also brings about some limitations, as the system's design was at the center of the discussion. Therefore, prominent features of the system, such as the solution space, may have prompted more discussion around topics such as shared understanding than other systems would. This means that the importance of some themes may be overemphasized, while other themes, which may be just as important, could be neglected. Perhaps an example of this could be seen in the topic of shared materials, as discussed previously. If CrowdMuse provided some support for it, there would likely be more discussion on the topic.

Finally, the data analysis was performed only by me. Therefore, there could be issues with the reliability of the coding, and my views may have biased the identification of the prominent themes in the discussion.

Conclusion

Large scale creativity research has made significant advances in improving ideation performance, usually by developing online brainstorming support systems. These systems, therefore, could be deployed to benefit online communities that currently engage in creative collaborations. However, it is unclear if the designs of current systems are adequate for such communities, as they are often developed with other goals in mind. To begin exploring this gap, I deployed CrowdMuse, a large-scale brainstorming system, into the online community for the FreeCAD open source project. Through an analysis of logs, posts in the discussion thread, and interviews, I described the behavior of users within the system, as well as their perceptions of the benefits and issues it could bring.

The results are informative for future research and designs for large-scale creativity. They point to 1) a different profile of ideas than those usually generated by crowd markets 2) the need for a more explicit form of support for divergence, and 3) the need to explore an integration of ideation support with other phases of the creative process. Despite some generalization limitations due to the small and homogenous sample, this chapter contributes with an exploration of points that should be attended to by the large-scale creativity community in order to improve the applicability of their findings to existing scenarios of online creative collaborations.

CHAPTER 7

FINAL DISCUSSION

In this dissertation, I explored the broad theme of large-scale creativity support. The overarching research question was "**How can we appropriately support large-scale creative collaborations in distributed online communities?**" This question was approached by looking at both the current practices in creative communities such as those behind open source projects, as well as advancing the support techniques that have been examined by the research community. In this chapter, I briefly review each chapter, and discuss their implications for research on crowd creativity. I end with a discussion of the future work this dissertation enables.

The first topic I approached was that of the adequacy of common communication channels for supporting the creativity of online communities (Chapter 3). This inquiry is motivated by the development of design principles for Creativity Support Tools (CST) in the past decades. Through a literature review, I consolidated a list of 6 heuristics that can determine a system's adequacy in supporting creativity. Using these heuristics as a framework, I analyzed the adequacy of the communication channels used by open source software projects through three methods: 1) a heuristic analysis of their channel's UIs; 2) a content analysis of the discussions; and 3) interview with open source participants. This analysis uncovered a list of creativity bottlenecks likely caused by these communication channels, such as favoring discussion over exploration, lack of support for iteration, as well as barriers for users with diverse background. In this chapter, I made the following contributions:

- A set of heuristics, grounded on the CST literature, that can be used to understand the appropriateness of software for supporting creativity;
- An understanding of visible breakdowns in the creativity of these groups, as well as the possible UI elements in their communication channels that may cause them.

In the following chapter (Chapter 4), I departed from evaluating current practices in online communities and looked into improving current support techniques in crowd creativity research. By looking at the present literature, I identified an opportunity for using the peripheral microtasks—tasks commonly done by a third-party crowd—as a way of improving the attention to the inspirations. I examined this hypothesis through 4 iterative studies on Mechanical Turk. The studies compared tasks to simple exposure or no inspirations, and varied in their task types, duration, number of ideas per inspiration, and selection mechanism. The results pointed to combination and rating tasks outperforming simple exposure, but mainly on the second half of ideation, and for high fluency ideators. They also pointed to the importance of the number of ideas per inspiration, the effects that idea selection mechanisms play. This resulted in the following contributions:

- Evaluation of the benefits that embedding tasks into inspirations brings to ideation performance, including which tasks performed better;
- An exploration of how different factors affect inspiration, such as the number of ideas per inspiration and their timing during ideation.

After that, I investigated a novel form of inspiration support (Chapter 5). Unlike the previous chapter, in which I attempted to extend the effect of existing approaches, in this chapter I proposed a first-of-its-kind adaptive brainstorming support system--CrowdMuse. CrowdMuse models ideators based on their ideation performance, keeping track of the categories of ideas individuals have visited. Using theoretical models of idea generation and recommender system techniques, the system adapts its visualization and inspiration mechanism to prioritize idea categories that have a higher potential of inspiring ideators. I evaluated the system's effects through two large-scale studies in Prolific. The results indicate that the inspiration mechanism was the only one able to affect performance—no results stemmed from the adaptive solution space. Furthermore, the effects were seen in breadth of ideation, suggesting an improvement brought about by the recommended categories. Finally, results also point to the need to find the right categorization level, as broad categories may not be enough to elicit any effects, while too fine categories may yield usability issues. The contributions from this chapter include:

- CrowdMuse, its modeling techniques, and adaptive mechanism;
- An evaluation of CrowdMuse's adaptive features on brainstorming

Finally, in Chapter 6, I began to join the two threads of research that were the foci of the preceding chapters. This was done by deploying CrowdMuse into the FreeCAD project community. The goal was to perform an initial exploratory analysis of CrowdMuse's adequacy as an alternative to the common communication channels used for creative collaborations. The deployment lasted for two weeks. Logs for the system provide a snapshot of users' behaviors. Interviews and discussions in the forum yield further information of users' perceptions of the system. The data points to three trends: 1)

While participants in crowd markets usually generate many short ideas, users in specialized communities may favor few better developed ideas; 2) the need to explore more structure mechanisms to induce divergence; and 3) The need for research to explore how to extend creativity support to other phases of creativity rather than only idea generation. This work brings implications for the design of large-scale creativity tools that are meant to be used by existing online communities. Contributions include:

- Indications of important differences in behavior between crowd market users and members in communities;
- Evidence of the need for research to explore support mechanisms to other phases of creativity in conjunction with idea generation;

Discussion

The first main contribution I intended to make towards crowd creativity research was that of using microtasks to boost attention to the inspirations and consequently their effects. The studies in Chapter 4 provided some evidence for that, showing particularly that combination and rating tasks were effective, under some conditions, in boosting attention. However, Chapter 5 also sheds further light into that. While the main purpose for CrowdMuse was to evaluate the effects of adaptations, it did so through two distinct channels: the inspiration mechanism and the solution space. I hypothesized that a greater effect would be seen by the adaptive inspiration mechanism compared to an adaptive solution space due to two of the mechanisms examined in Chapter 4: increased attention due to microtasks, as well as a moderate number of ideas presented at once (3). The results were clear: positive effects on breadth were found on participants who were presented with an adaptive inspiration mechanism but not the solution space. Although the latter may have suffered due to some usability issues, users still felt it was reasonably useful, and if fared rather close to the inspiration mechanism in the final questionnaire in study 1 of Chapter 5. In that questionnaire, users answered the question "*How useful was the solution space*" with an average of 5.12 out of 7, compared to an average of 5.30 for the question "*How useful was the inspiration mechanism*". Therefore, I argue that both Chapters provide strength to my hypothesis of microtasks drawing further attention to inspirations, and therefore magnifying their effects.

While this microtask-based attention mechanism was incremental to existing approaches in the literature, the adaptive mechanisms presented in Chapter 5 represent a completely new direction for investigation. The modelling and adaptions presented in that chapter were designed mainly after two well-accepted models of idea generation, primarily SIAM (Nijstad & Stroebe, 2006), but with contributions from the Matrix Model (Brown et al., 1998). To very briefly review each model (see Chapter 2 for more information), SIAM defines idea generation happening around two loops—category retrieval, followed by idea generation within the category. The Matrix Model proposes that category switching is a probabilistic mechanism, and probabilities vary based on current and next categories. By keeping track of the current category, CrowdMuse should be able to extend the idea generation loop. By keeping track of adjacent categories and their frequencies, CrowdMuse should be able to make more informed guesses of the probability for revisiting previous categories and therefore extend the category loading loop. Finally, by inferring recommended categories, CrowdMuse should be able to extend the category loading loop. The results presented in Chapter 5 indicate that CrowdMuse was only effective at the category loading loop, and only for loading new categories.

The important question to ask then is *why*? Before approaching possible answers, it is important to consider a few of other factors. The first is that participants showed an extremely high inclination towards category switching (e.g. participants moved to different categories 95% of the time in study 1). Another interesting factor to consider is evident by a look at the literature. We find works based on idea exposure, particularly those that favour diversity of ideas, having an effect on the breadth or diversity of ideas. This is seen in the two chapters of this dissertation as well as the work from Siangliulue Arnold, et al. (2015). On the other hand, the work from Chan et al. (2016b) found improvements on depth and fluency when the session was facilitated by skilled facilitators. Breadth did not improve. Their qualitative analysis of facilitation strategies showed that skilled facilitators favoured inspirations based on simulations (e.g. "Imagine yourself in this situation…") rather than simple exposure to other ideas. A final factor for consideration is seen in Study 2B of Chapter 4. In that study, I controlled the inspiration mechanism's pool of ideas in order for it to always show related ideas. The results show that participants in the microtask condition actually performed worse than those in the idea exposure condition.

A very interesting implication that arises from these factors is that different inspiration mechanisms may be needed for extending each of the two loops in the SIAM model. Idea exposure may be enough to expand the category loading loop to new categories, particularly if exposure is accompanied by some attention boosting mechanisms such as microtasks. However, it may not be enough to extend the idea generation loop, either through the current or previously visited categories. Study 2B even provides some evidence that exposure could hinder that loop. But given the evidence shown by Chan et al.'s facilitation study, it seems that simulation questions could be one way of expanding that loop. Therefore, CrowdMuse could have been successful in supporting current or previously visited categories by promoting simulations rather than exposure. This would mean that when a user clicked the inspiration mechanism, instead of showing other ideas it could, for example, show "*what if*" questions based on their previous ideas. But this may also open further questions related to adaptation, possibly resulting in systems that adapt not only based on categories, but also on cognitive strategies. For example, are different inspiration strategies more or less useful to different ideators? That is, would two different ideators benefit equally from simulation questions? Or are they more useful early on ideation or later? Such questions yield abundant new lines of inquiry that could make important contributions to theoretical idea generation models as well as adaptive creativity system designs.

The categorization scheme is clearly of core importance to discussions of adaptation, since it is what informs CrowdMuse's models and adaptations. The studies showed positive results with over 40 categories mostly developed by the users themselves, and null results when using a much smaller scheme generated by me and another coder. Therefore, there are two factors that could be at play here: the closeness of the taxonomy to the ideators' expectations and level of category detail. The first may be particularly challenging. Could it be that a categorization developed by researchers would be less useful than one developed by communities themselves? Another consideration is that of cultural differences in categorization. Ji, Zhang, & Nisbett (2004) have shown cultural and language effects on categorization, with Americans favouring grouping based on taxonomies and East Asians favouring grouping based on themes. Would this affect the core categorization model of ideas in CrowdMuse and its adaptive mechanisms? These questions are particularly important in both crowd markets and open source projects, as they present very diverse geographical compositions. As for the level of detail of the categories, the issue here is to strike a balance between a scheme that is general enough to be usable, and fine-grained enough to be useful. As discussed in Chapter 5, the studies in this dissertation only provide some initial indications of the important factors and their effects but is limited in making deeper inferences about them. Therefore, future work must systematically explore these issues to understand their effects and applicability to different contexts.

Another important point of discussion is on whether the inspiration approach generally used in Chapters 4 and 5 may be inducive to priming or fixation effects. Research has shown that seeing examples may increase conformity of new ideas to the examples, although not in ways that constrain creative output (Marsh, Landau, & Hicks, 1996). In the studies presented in this dissertation, the one possible evidence of fixation happened in study 2B (Chapter 4), in which the inspiration mechanism presented closely related ideas and asked users to judge the similarity of the ideas. This may have caused users to focus too much on the details of similar ideas, which can lead to fixation (Chan et al., 2016a; Jansson & Smith, 1991). Other than that, not much evidence was seen. For example, the practice of almost always alternating to different categories seen in the CrowdMuse studies indicates that the examples did not overly prime users into only a few overall topics. However, changes could be introduced to further minimize the risks of priming and fixation effects. For example, the system could only enable the inspiration mechanism after the user has already entered a few ideas. This would prevent a very early priming from external stimuli, although the users' own initial ideas could still cause the priming to occur.

Finally, I briefly discuss the issue of generalization of crowd platform results to other populations and contexts. This research followed common practices in the crowd creativity literature. Studies were performed on two different crowd markets (Mechanical Turk and Prolific) and evaluated some form of intervention to improve brainstorming process metrics such as fluency, breadth, and depth. This may prompt questions on our ability to generalize these results to other contexts, particularly those outside of crowd markets. To improve this quality, I specifically chose problems that would be relevant for each community. For Mechanical Turk it was the development of a Mechanical Turk mobile app (proposed by Krynicki, 2014, and also used by Chan et al., 2016b), and for Prolific it was ideas to improve the service. Both of these were chosen to maximize the likelihood of users being knowledgeable and motivated to brainstorm (T. Amabile, 1983). For example, in Chapter 5's Study 3, users reported being reasonably attentive (~3.5 out of 5) and involved (~4.1 out of 5) with the task based on their survey responses. In Chapter 6's Study 1, Prolific users reported being motivated (~5.3 out of 7) and knowledgeable (~5.4 out of 7) to generate ideas to improve the service. Consequently, they have the potential to be creative. Furthermore, the results are in accordance with the body of creativity research, a large part of which has been conducted in a different context than the one in this dissertation. Therefore, the results obtained through these crowd markets should be an appropriate indication of those effects in similar populations and contexts. On the other hand, the deployment provided some information of when these results may not be immediately applicable, as will be discussed soon.

The other part of this dissertation focused on the broader topic of creativity support in existing online communities. This was carried out through the bottleneck analysis in Chapter 3 and the deployment in Chapter 6. I particularly focused on Open Source projects. Open source contributors have been said to show a high level of motivation and expertise, often experiencing a high sense of creativity and flow while contributing to open source communities (Lakhani & Wolf, 2003). Furthermore, such collaborations have tangible results in popular projects such as Linux, Apache, GIMP, and many others. Their participation is also voluntary, avoiding issues with extrinsic financial motivations (T. M. Amabile, 1985).

All of these factors allow us to enrich the discussion on crowd market studies usually found in the literature. The first point is that of differences in idea generation strategies and behaviours between crowd workers and open source contributors. As discussed in Chapter 6, FreeCAD users demonstrated a different approach to ideation on CrowdMuse when compared to crowd market participants. In crowd markets, users have a single quick interaction in which they try to come up with as many ideas as possible. In a deployment such as the one done in FreeCAD, idea generators have more time flexibility and the opportunity to return later. Additionally, each FreeCAD ideator only added about two ideas but they were much more detailed. So, due to a different strategy, the approaches commonly used in crowd markets (including those evaluated in Chapters 4 and 5) may not be immediately effective. These communities may need a greater structure if interventions commonly used in the crowd creativity literature are to be used. For example, CrowdMuse could be more explicit about the expected process, such as by limiting characters and suggesting a target number of ideas before allowing users to expand on the ideas. Microtask crowdsourcing research can be informative in how to break this process down (Kittur et al., 2011; Teevan, Iqbal, & von Veh, 2016b).

Another interesting distinction pertains to the solution space. As previously discussed, the solution space did not cause differences in idea generation performance. But this does not mean that it should be abandoned, especially in light of the deployment in FreeCAD. The solution space was valued by the FreeCAD users who used the system specifically for the reason I had hypothesized in Chapter 5: it gives the users the initiative to explore the solution space, contrary to the inspiration mechanism. This points to a need to balance freedom and exploration with information overload and system control. Of course, the small, admin-biased, open source sample prevents us from making conclusive claims, but this contrast points to an interesting characteristic of open source contributors that contrasts with crowd workers.

Temporal aspects are also not considered in common crowd studies. They expect users to use the system once for a pre-determined amount of time. In the FreeCAD deployment, users did not have that requirement, and could return several times. Both the analyses done in Chapter 3 as well as the deployment picked up on such issues with discussion boards. Users had a hard time going back to a topic, as that would imply reading all the messages posted since their last visit. Therefore, the contextual work presented in this dissertation points to the need of further research on summarization in asynchronous brainstorming sessions. Earlier work from Farooq et al. (2007) began to look into this by suggesting a log-based summarization and task suggestion feature. Crowd creativity research could investigate how scale such mechanisms.

The deployment study presented in Chapter 6 lends further support to the findings of chapter 3. The interviews with FreeCAD users surfaced many of the topics I discussed in the bottleneck analysis, such as ideas being lost in discussion, unproductive discussions, early convergence, users not reading previous messages and having to manually update others of any progress, and barriers to participation of all users, many of these explicitly associated by the users to the discussion forum structure. This strengthens the notion of creativity bottlenecks caused by UI design, as well as the usefulness of the set of heuristics for evaluating adequacy of systems for creative collaborations. On a final anecdotal note, I highlight the importance that having CrowdMuse as the center of conversations brought to interviews with open source participants. When I conducted the interviews for Chapter 3, it was difficult grounding users' notions of creativity and the harm that traditional communication channels could inflict on it. Participants did not demonstrate much issue with their processes and tools. In contrast, the interviews conducted in Chapter 6 exposed those issues much more clearly. Users easily discussed their perceived advantages to a system such as CrowdMuse and deliberated on the issues they faced with the discussion forum, particularly for idea generation. This could indicate users' lack of awareness about creativity processes and bottlenecks, which further stresses the need for large-scale creativity researchers to understand the populations they intend on benefitting instead of limiting such research to crowd markets.

I end this section with a brief discussion on two broader points related to this research: 1) the benefits of crowd creativity as opposed to a small groups or individuals;

2) the characteristics of crowds and domains that can benefit from the approach investigated in this dissertation.

What benefits does a crowd approach for creativity bring over traditional individual or small group creativity? I argue that the greatest strengths for a large-scale approach are in the large amount of data it outputs as well as its low barrier for entry. The amount of data enables, for example, the adaptive features of CrowdMuse. Alternatively, crowds can be used to systematically explore the solution space (Huang & Quinn, 2017). As for barrier for entries, a large-scale approach will likely aim to reach great number of users, regardless of their geographical location, experience, credentials, or belonging to certain groups of communities. This is in agreement with the argument proposed by Fischer (2005), who claims that the removal of such barriers can foster creativity. Participation is decided solely by knowledge and interest, not by other external constraints. But large-scale creativity is not yet at this point. For example, some of the communities evaluated in this work also invest in gathering its members in one place through conferences or through smaller meetings. This provides some evidence of the insufficiency of their online communication processes. Therefore, as argued throughout this work, more research is needed into the entire process of large-scale creativity (rather than only idea generation), especially into supporting convergent and decision-making processes. Without this broader work, the applicability or usefulness of large-scale creativity may remain limited.

The second and final point is: are there specific kinds of crowds and domains to whom this approach can be particularly useful? The microtask interventions were designed for crowds that have limited time availability (such as MTurk or FOSS communities). They will, therefore, have only a quick interaction with the brainstorming system. In this context, increasing attention to the important ideas is key to maximize effect during this short time. In a context in which users are expected to spend greater time exploring the ideas (such as a professional one), attention boosting interventions may not be as important. Furthermore, these interventions are designed for groups that can generate a large number of ideas but do not have the resources to appropriately deal with all of them. In this sense, the adaptive approach employed here works as a filter to direct ideators to what should be more inspiring. The underlying assumption, therefore, is that users can be clustered into diverse groups that can maximize inspiration amongst themselves. Therefore, it is important for crowds to be diverse enough in order to supply enough inspiration within each cluster. In summary, the approach investigated here is particularly useful for crowds that: 1) are limited in their time and ability to deal with a large amount of ideas; 2) are diverse enough to enable clusters of users to generate enough inspiration content for themselves.

The final point is about which domains lend themselves to the approach explored in this dissertation. This issue has not yet been explored in depth. This is likely due to most of the research being done in crowd markets, which severely constrains the domain options. However, in this dissertation I began to explore idea generation outside of crowd markets. I already discussed the differences in behaviors between crowd markets and FOSS communities, attributing them to the characteristics of the users and communities. However, it is possible that some of that may be due to the domain of the ideation prompt. The ideation sessions in Chapter 5 targeted ideas to improve the Prolific service. As the resulting categories indicate, there are many different areas of focus for doing that, such as improving payment and access to research results. Users are likely to be fluent in these categories, since they generally are Prolific features with which they interact daily. However, it may be hard to remember them all, and even more to generate ideas in them. In this context, CrowdMuse's features can be useful for bringing (useful) categories to mind while also presenting example ideas. In summary, the domain explored in Chapter 5 had the following characteristics: 1) many different solution categories; 2) the categories are fairly easy to identify; and 3) the ideators are likely to be fluent on many of these categories, even if they are not immediately accessible to them. On the other hand, the problem explored in Chapter 6 (improving FreeCAD adoption in education) may not have fulfilled these requirements. While there may be many ways of achieving that result, it is not immediately obvious what those categories are. Furthermore, it is possible that not many people would have enough fluency within each of those categories. Therefore, the domain explored in Chapter 6 violated requirements 2 and 3. However, these requirements are only suggestions based on the studies here presented, especially given the exploratory nature of Chapter 6. Future work could seek to validate or expand these requirements.

Limitations

I now briefly reiterate some of the limitations of this work. They have already been discussed in more details in each chapter, but I summarize them here for both strands of this dissertation. The work on improving inspiration support through microtasks does not examine the output from the microtasks, and the feasibility of turning them into something useful for the ideation process (e.g. a consistent rating of the ideas).

As for CrowdMuse, it points to the importance of the categorization scheme, including some important features such as the granularity level. However, it does not explore this important issue in more detail. Furthermore, the system was evaluated with an existing pool of ideas and users, which means that I did not explore dealing with the issue of cold start. Finally, both chapters also have some metric limitations, such as not using any direct metric for evaluating the creativity of ideas.

As for the work done in online communities, the main issue was that the only context analyzed in this work was that of FOSS communities. Therefore, it is unclear to what extent those results generalize to other kinds of online communities. Furthermore, there were issues with low participation, as well as considerations in relation to the coding strategies used. For example, the lack of reliability metrics between different coders may cloud the extent to which the results could be reliably reproduced.

Future Work

The discussion in this chapter has already outlined some avenues for future work. Naturally, there is some incremental research that could and should be done to strengthen and extend the research discussed here. For example, the conclusions relating to the deployment of CrowdMuse would greatly benefit from recruiting many other communities, in different domains, and with different orientations (e.g. communities of interest). But there are also high-impact topics that this research suggests. Therefore, I now describe four topics I believe to be among the most promising and interesting research paths suggested by this work.

Supporting Creativity through Adaptive Strategies

In this work, I explored adaptations around categories of ideas. However, as discussed above, it is possible that different parts of the cognitive processes of idea generation have to be supported by different strategies. Previous research also suggests that individual ideators can differ in their approaches. For example, previous work distinguishes between ideators with a divergent and convergent orientations (Nijstad et al., 2010). Further differences were also seen in the contrasts between crowd market and open source ideators (Chapter 6). Therefore, future research should investigate ways to make systems adapt not only to ideation categories, but also to strategies. Example questions include: what is the space of possible ideation strategies? How do different strategies affect idea generation? Do ideators change their strategies over time? Can we infer ideator strategies? Exploring these questions brings implications not only to the development of creativity support systems, but also to the theoretical models of creativity.

Supporting the Creative Process Entirely

As discussed previously, crowd creativity research (or even the broader field of creativity) has greatly focused its efforts towards idea generation and fostering divergence, usually neglecting other phases of creativity. This was particularly clear based on FreeCAD members comments on CrowdMuse. Similar systems would likely have elicited similar comments. Therefore, it is critical for crowdsourced creativity to focus on supporting creativity during other phases of the creative process in addition to idea generation. For example, how long should the idea generation session last in a crowd context? Can users go through different phases asynchronously or should the entire community be synced to the same phase? Should the process follow linearly, or should it be cyclical? How to ensure participation throughout the entire process, or how to deal with attrition? How to conciliate individual creativity processes to the crowd's?

Problem-agnostic Inspirations

The recommender system techniques used in CrowdMuse proved to be useful in inferring new categories for idea generation. However, this means that CrowdMuse also shares the same issue with recommender systems: the cold start problem (Lü et al., 2012). The problem happens when there is no or not enough data in the system in order to make inferences. This was the situation of the FreeCAD deployment. The first participant to use CrowdMuse found an empty solution space, and therefore no support. Even when a few ideas were already added, making such inferences is not that useful, as users can quickly browse all the ideas so far. A possible solution for this issue could lie in the development of a pool of problem-agnostic set of inspirations. Earlier work has already started to probe into this concept (Yu et al., 2014b, 2014a; Yu, Kraut, et al., 2016). Research questions could include: how could these inspirations be first generated? how can a set of general inspirations be chosen to populate a new problem? How can these inspirations be categorized for each different problem?

Problem Finding and Cross-Community Creativity

In such a context, recommender system techniques could be used not only to infer new categories within a problem, but also to infer useful new problems an ideator could be knowledgeable and motivated to participate in. A broader pool of users could increase the potential for creativity. If problems could be recommended and access across community lines, it could increase the diversity and knowledge of the ideation group. For example, members of other open source communities used in educational settings could have been alerted to the ongoing brainstorm session on increasing FreeCAD's adoption in education, and both communities could collaborate with ideas towards their common goals of improving adoption in education. A shared pool of brainstorming sessions could also contribute towards overcoming the cold start problem.

Conclusion

In this dissertation, I explored ways of improving support mechanisms in crowd idea generation, as well as the creativity issues and needs in open source communities. The combined results point to ways of improving attention to inspirations thus increasing their effects; to the potential for adaptive mechanisms to increase idea generation performance; to the issues currently faced by online communities in their creative collaborations; and to how a crowd idea generation system could help an online community, as well as the areas where it still falls short. Therefore, this work contributes to the growing body of literature on supporting large-scale creativity with incremental and novel techniques. It also encourages future research to focus on populations and creativity phases that remain largely untouched.

REFERENCES

Alsallakh, B., Micallef, L., Aigner, W., Hauser, H., Miksch, S., & Rodgers, P. (2016). The State-of-the-Art of Set Visualization. *Computer Graphics Forum*, *35*(1), 234–260. https://doi.org/10.1111/cgf.12722

Amabile, T. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357.

Amabile, T. (1988). A model of creativity and innovation in organizations. *Research in Organizational Behavior*, *10*(1), 123–167.

Amabile, T. (2012). *Componential theory of creativity*. Harvard Business School. Retrieved from http://www.hbs.edu/faculty/Publication%20Files/12-096.pdf

Amabile, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology*, *48*(2), 393.

Amabile, T., & Mueller, J. (2002). Assessing creativity and its antecedents: An exploration of the componential theory of creativity. In C. Ford & N. Mahwah (Eds.), *Handbook of organizational creativity* (pp. 33–64). London: Lawrence Erlbaum Associates.

Arias, E., Eden, H., Fischer, G., Gorman, A., & Scharff, E. (2000). Transcending the individual human mind---creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction*, *7*(1), 84–113. https://doi.org/10.1145/344949.345015

Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., … Panovich, K. (2010). Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (pp. 313–322). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1866078

Bigham, J. P., Bernstein, M., & Adar, E. (2015). Human-Computer Interaction and Collective Intelligence. In *Handbook of Collective Intelligence*.

Bittner, E. A. C., & Leimeister, J. M. (2014). Creating Shared Understanding in Heterogeneous Work Groups: Why It Matters and How to Achieve It. *Journal of Management Information Systems*, *31*(1), 111–144. https://doi.org/10.2753/MIS0742-1222310106

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

Brown, V., Tumeo, M., Larey, T. S., & Paulus, P. B. (1998). Modeling Cognitive Interactions During Group Brainstorming. *Small Group Research*, *29*(4), 495–526.

Chan, J., Dang, S., & Dow, S. P. (2016a). Comparing Different Sensemaking Approaches for Large-Scale Ideation. Presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Retrieved from http://joelchan.me/files/2016-chi-sensemaking-ideation.pdf

Chan, J., Dang, S., & Dow, S. P. (2016b). Improving Crowd Innovation with Expert Facilitation. Presented at the Computer Supported Collaborative Work.

Chan, J., & Schunn, C. D. (2015). The importance of iteration in creative conceptual combination. *Cognition*, *145*, 104–115. https://doi.org/10.1016/j.cognition.2015.08.008

Cheng, J., Teevan, J., Iqbal, S. T., & Bernstein, M. S. (2015). Break It Down: A Comparison of Macro- and Microtasks (pp. 4061–4064). ACM Press. https://doi.org/10.1145/2702123.2702146

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chilton, L. B., Little, G., Edge, D., Weld, D. S., & Landay, J. A. (2013). Cascade: Crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1999–2008). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2466265

Coskun, H., Paulus, P. B., Brown, V., & Sherwood, J. J. (2000). Cognitive stimulation and problem presentation in idea-generating groups. *Group Dynamics: Theory, Research, and Practice*, *4*(4), 307–329. https://doi.org/10.1037/1089-2699.4.4.307

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, *18*(3), 391–404.

Cropley, A., & Cropley, D. (2008). Resolving the paradoxes of creativity: an extended phase model. *Cambridge Journal of Education*, *38*(3), 355–373. https://doi.org/10.1080/03057640802286871

Csikszentmihalyi, M. (1997). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: Harper Perennial.

Csikszentmihalyi, M. (1999). Implications of a Systems Perspective for the Study of Creativity. In *Handbook of Creativity*.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: transparency and collaboration in an open software repository (p. 1277). ACM Press. https://doi.org/10.1145/2145204.2145396

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

Davis, N., Zook, A., O'Neill, B., Headrick, B., Riedl, M., Grosz, A., & Nitsche, M. (2013). Creativity support for novice digital filmmaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 651–660). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2470747

Dennis, A. R., & Williams, M. L. (2003). Electronic Brainstorming: Theory, Research, and Future Directions. In *Group creativity: Innovation through collaboration*. Oxford University Press.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509. https://doi.org/10.1037/0022-3514.53.3.497

Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. *Journal of Personality and Social Psychology*, *61*(3), 392.

Doboli, A., Umbarkar, A., Subramanian, V., & Doboli, S. (2014). Two experimental studies on creative concept combinations in modular design of electronic embedded systems. *Design Studies*, *35*(1), 80–109. https://doi.org/10.1016/j.destud.2013.10.002

Dollinger, S. J., & Shafran, M. (2005). Note on Consensual Assessment Technique in Creativity Research. *Perceptual and Motor Skills*, *100*(3), 592–598. https://doi.org/10.2466/pms.100.3.592-598

Dow, S., Kulkarni, A., Klemmer, S., & Hartmann, B. (2012). Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1013–1022). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2145355

Dow, S. P., Heddleston, K., & Klemmer, S. R. (2009). The efficacy of prototyping under time constraints. In *Proceedings of the seventh ACM conference on Creativity and cognition* (pp. 165–174). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1640260

Dugosh, K. L., Paulus, P. B., Roland, E. J., & Yang, H.-C. (2000). Cognitive stimulation in brainstorming. *Journal of Personality and Social Psychology*, *79*(5), 722–735. https://doi.org/10.1037/0022-3514.79.5.722

Estelles-Arolas, E., & Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189–200. https://doi.org/10.1177/0165551512437638

Farooq, U., Carroll, J. M., & Ganoe, C. (2008). Designing for creativity in computer-supported cooperative work.

Farooq, U., Carroll, J. M., & Ganoe, C. H. (2007). Supporting creativity with awareness in distributed collaboration. In *Proceedings of the 2007 international ACM conference on Supporting group work* (pp. 31–40). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1316630

Fischer, G. (2005). Distances and diversity: sources for social creativity (p. 128). ACM Press. https://doi.org/10.1145/1056224.1056243

Gabriel, A., Monticolo, D., Camargo, M., & Bourgault, M. (2016). Creativity support systems: A systematic mapping study. *Thinking Skills and Creativity*, *21*, 109–122. https://doi.org/10.1016/j.tsc.2016.05.009

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*(7070), 900–901.

Girotto, V., Walker, E., & Burleson, W. (2017). The Effect of Peripheral Micro-tasks on Crowd Ideation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM.

Girotto, V., Walker, E., & Burleson, W. (2018). CrowdMuse: An Adaptive Crowd Brainstorming System. In *The 31st Annual ACM Symposium on User Interface*

Software and Technology Adjunct Proceedings - UIST '18 Adjunct (pp. 93–95). Berlin, Germany: ACM Press. https://doi.org/10.1145/3266037.3266112

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*, 444–454.

Guilford, J. P., Merrifield, P. R., & Wilson, R. C. (1958). *Unusual Uses Test*. Orange, CA: Sheridan Psychological Services.

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology*, *61*(1), 569–598. https://doi.org/10.1146/annurev.psych.093008.100416

Hennessey, B. A., Amabile, T. M., & Mueller, J. (2011). Consensual Assessment. In *Encyclopedia of Creativity* (2nd ed., pp. 253–260). San Diego: Academic Press.

Herrmann, T. (2009). Design heuristics for computer supported collaborative creativity. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* (pp. 1–10). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4755350

Hewett, T. T. (2005). Informing the design of computer-based environments to support creativity. *International Journal of Human-Computer Studies*, *63*(4–5), 383–409. https://doi.org/10.1016/j.ijhcs.2005.04.004

Hinds, P. J., & Weisband, S. P. (2003). Knowledge sharing and shared understanding in virtual teams. *Virtual Teams That Work: Creating Conditions for Virtual Team Effectiveness*, 21–36.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14*(6), 1–4.

Huang, G., & Quinn, A. J. (2017). BlueSky: Crowd-Powered Uniform Sampling of Idea Spaces (pp. 119–130). ACM Press. https://doi.org/10.1145/3059454.3059481

Ivanov, A., & Cyr, D. (2006). The Concept Plot: a concept mapping visualization tool for asynchronous web-based brainstorming sessions. *Information Visualization*, *5*(3), 185–191. https://doi.org/10.1057/palgrave.ivs.9500130

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

Ji, L.-J., Zhang, Z., & Nisbett, R. E. (2004). Is It Culture or Is It Language? Examination of Language Effects in Cross-Cultural Research on Categorization. *Journal of*

Personality and Social Psychology, *87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57

Kanefsky, B., Barlow, N. G., & Gulick, V. C. (2001). Can distributed volunteers accomplish massive data analysis tasks. *Lunar and Planetary Science*, *1*. Retrieved from http://www.nasaclickworkers.com/classic/documents/abstract.pdf

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology*, *13*(1), 1–12. https://doi.org/10.1037/a0013688

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 453–456). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1357127

Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011). Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 43–52). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2047202

Kohn, N. W., Paulus, P. B., & Choi, Y. (2011). Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology*, *47*(3), 554–561. https://doi.org/10.1016/j.jesp.2011.01.004

Kozbelt, A., Beghetto, R. A., & Runco, M. A. (2010). Theories of creativity. *The Cambridge Handbook of Creativity*, 20–47.

Krynicki, F. (2014). *Methods and models for the quantitative analysis of crowd brainstorming*. Retrieved from https://uwspace.uwaterloo.ca/handle/10012/8347

Lakhani, K., & Wolf, R. G. (2003). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.443040

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284.

Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2010). TurKit: human computation algorithms on mechanical turk (p. 57). ACM Press. https://doi.org/10.1145/1866029.1866040

Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics Reports*, *519*(1), 1–49. https://doi.org/10.1016/j.physrep.2012.02.006

Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the Special Issue. *International Journal of Human-Computer Studies*, *63*(4–5), 365–369. https://doi.org/10.1016/j.ijhcs.2005.04.002

Luther, K., Tolentino, J.-L., Wu, W., Pavel, A., Bailey, B. P., Agrawala, M., … Dow, S. P. (2015). Structuring, Aggregating, and Evaluating Crowdsourced Design Critique (pp. 473–485). ACM Press. https://doi.org/10.1145/2675133.2675283

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

McDonald, N., & Goggins, S. (2013). Performance and participation in open source software on GitHub (p. 139). ACM Press. https://doi.org/10.1145/2468356.2468382

McInnis, B. J., Murnane, E. L., Epstein, D., Cosley, D., & Leshed, G. (2016). One and Done: Factors affecting one-time contributors to ad-hoc online communities (pp. 608–622). ACM Press. https://doi.org/10.1145/2818048.2820075

McLeod, P. L., Lobel, S. A., & Cox Jr., T. H. (1996). Ethnic Diversity and Creativity in Small Groups. *Small Group Research*, *27*(2), 248–264.

Milliken, F. J., Bartel, C. A., & Kurtzberg, T. R. (2003). Diversity and Creativity in Work Groups: A Dynamic Perspective on the Affective and Cognitive Processes That Link Diversity and Performance. In *Group creativity: Innovation through collaboration*. Oxford University Press.

Nakakoji, K. (2006). Meanings of tools, support, and uses for creative design processes. In *International Design Research Symposium ‘06, Seoul, Korea*. Retrieved from http://www.kidresearch.jp/publications/IDRS-nakakoji-rev-wt-header.pdf

Nelson, B. A., Wilson, J. O., Rosen, D., & Yen, J. (2009). Refined metrics for measuring ideation effectiveness. *Design Studies*, *30*(6), 737–743. https://doi.org/10.1016/j.destud.2009.07.002

Nemeth, Charlan J., Personnaz, B., Personnaz, M., & Goncalo, J. A. (2004). The liberating role of conflict in group creativity: A study in two countries. *European Journal of Social Psychology*, *34*(4), 365–374. https://doi.org/10.1002/ejsp.210

Nemeth, Charlan Jeanne, & Nemeth-Brown, B. (2003). Better than individuals. In *Group creativity: Innovation through collaboration* (pp. 63–84).

Netcraft. (2018, April 15). January 2018 Web Server Survey. Retrieved April 15, 2018, from http://news.netcraft.com/archives/2015/09/16/september-2015-web-server-survey.html

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces (pp. 249–256). ACM Press. https://doi.org/10.1145/97243.97281

Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, *21*(1), 34–77. https://doi.org/10.1080/10463281003765323

Nijstad, B. A., Diehl, M., & Stroebe, W. (2003). Cognitive Stimulation and Interference in Idea-Generating Groups. In *Group Creativity: Innovation Through Collaboration*. Oxford University Press.

Nijstad, B. A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, *10*(3), 186–213.

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

Olteţeanu, A.-M., & Falomir, Z. (2016). Object replacement and object composition in a creative cognitive system. Towards a computational solver of the Alternative Uses Test. *Cognitive Systems Research*, *39*, 15–32. https://doi.org/10.1016/j.cogsys.2015.12.011

Open Source Initiative. (2007). The Open Source Definition. Retrieved April 13, 2018, from https://opensource.org/osd

Osborn, A. F. (1963). *Applied imagination; principles and procedures of creative problem-solving* (Vol. 3d rev.). New York: Scribner.

Paulus, P. B., & Brown, V. R. (2003). Enhancing ideational creativity in groups: Lessons from research on brainstorming. In *Group creativity: Innovation through collaboration*. Oxford University Press.

Plucker, J. A., & Makel, M. C. (2010). Assessment of Creativity. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (pp. 48–73). Cambridge: Cambridge University Press. Retrieved from http://ebooks.cambridge.org/ref/id/CBO9780511763205A013

Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Raymond, E. S. (1999). *The Cathedral and the Bazaar. Musings on Linux and Open Source by an Accidental Revolutionary.* Beijing: O'Reilly Media.

Resnick, M., Myers, B., Nakakoji, K., Shneiderman, B., Pausch, R., Selker, T., & Eisenberg, M. (2005). Design principles for tools to support creative thinking. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1822&context=isr

Retelny, D., Robaszkiewicz, S., To, A., Lasecki, W. S., Patel, J., Rahmati, N., … Bernstein, M. S. (2014). Expert crowdsourcing with flash teams (pp. 75–85). ACM Press. https://doi.org/10.1145/2642918.2647409

Runco, M. A. (2008). Commentary: Divergent thinking is not synonymous with creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 93–96. https://doi.org/10.1037/1931-3896.2.2.93

Runco, M. A. (2014). Conclusion: What Creativity is and What it is Not. In *Creativity* (pp. 389–427). Elsevier. https://doi.org/10.1016/B978-0-12-410512-6.00013-8

Saeed, S., Rohde, M., & Wulf, V. (2011). Communicating in a Transnational Network of Social Activists: The Crucial Importance of Mailing List Usage. In A. S. Vivacqua, C. Gutwin, & M. R. S. Borges (Eds.), *Collaboration and Technology* (Vol. 6969, pp. 81–88). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23801-7_7

Selker, T. (2005). Fostering motivation and creativity for computer users. *International Journal of Human-Computer Studies*, *63*(4–5), 410–421. https://doi.org/10.1016/j.ijhcs.2005.04.005

Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, *24*(2), 111–134. https://doi.org/10.1016/S0142-694X(02)00034-0

Shneiderman, B., Fischer, G., Czerwinski, M., Resnick, M., Myers, B., Candy, L., … others. (2006). Creativity support tools: Report from a US National Science Foundation sponsored workshop. *International Journal of Human-Computer Interaction*, *20*(2), 61–77.

Siangliulue, P., Arnold, K. C., Gajos, K. Z., & Dow, S. P. (2015). Toward Collaborative Ideation at Scale: Leveraging Ideas from Others to Generate More Creative and Diverse Ideas (pp. 937–945). ACM Press. https://doi.org/10.1145/2675133.2675239

Siangliulue, P., Chan, J., Dow, S. P., & Gajos, K. Z. (2016). IdeaHound: Improving Large-scale Collaborative Ideation with Crowd-Powered Real-time Semantic Modeling (pp. 609–624). ACM Press. https://doi.org/10.1145/2984511.2984578

Siangliulue, P., Chan, J., Gajos, K. Z., & Dow, S. P. (2015). Providing Timely Examples Improves the Quantity and Quality of Generated Ideas (pp. 83–92). ACM Press. https://doi.org/10.1145/2757226.2757230

Smith, S. M. (2003). The constraining effects of initial ideas. In P. B. Paulus & B. A. Nijstad (Eds.), *Group creativity: Innovation through collaboration* (pp. 15–31). New York, NY, US: Oxford University Press.

Stasser, G. (1992). Information Salience and the Discovery of Hidden Profiles by Decision-Making Groups: A “Thought Experiment.” *Organizational Behavior and Human Decision Processes*, 156–181.

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In *Handbook of Creativity* (pp. 3–15).

Teevan, J., Iqbal, S. T., & von Veh, C. (2016a). Supporting Collaborative Writing with Microtasks (pp. 2657–2668). ACM Press. https://doi.org/10.1145/2858036.2858108

Teevan, J., Iqbal, S. T., & von Veh, C. (2016b). Supporting Collaborative Writing with Microtasks (pp. 2657–2668). ACM Press. https://doi.org/10.1145/2858036.2858108

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, *27*(2), 237–246. https://doi.org/10.1177/1098214005283748

Thomas, M. J. W. (2002). Learning within incoherent structures: the space of online discussion forums: Learning within incoherent structures. *Journal of Computer Assisted Learning*, *18*(3), 351–366. https://doi.org/10.1046/j.0266-4909.2002.03800.x

Torrance, E. P., Ball, O. E., & Safter, H. T. (2003). *Torrance Tests of Creative Thinking*.

Valentine, M. A., Retelny, D., To, A., Rahmati, N., Doshi, T., & Bernstein, M. S. (2017). Flash Organizations: Crowdsourcing Complex Work by Structuring Crowds As Organizations (pp. 3523–3537). ACM Press. https://doi.org/10.1145/3025453.3025811

Vasilescu, B., Filkov, V., & Serebrenik, A. (2015). Perceptions of Diversity on Git Hub: A User Survey (pp. 50–56). IEEE. https://doi.org/10.1109/CHASE.2015.14

Vasilescu, B., Serebrenik, A., Devanbu, P., & Filkov, V. (2014). How social Q&A sites are changing knowledge sharing in open source software communities (pp. 342–354). ACM Press. https://doi.org/10.1145/2531602.2531659

Wallas, G. (1926). *The Art of Thought*.

Wang, H.-C., Cosley, D., & Fussell, S. R. (2010). Idea Expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 103–106). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1718938

Wang, K., & Nickerson, J. V. (2017). A literature review on individual creativity support systems. *Computers in Human Behavior*, *74*, 139–151. https://doi.org/10.1016/j.chb.2017.04.035

Wang, Y.-C., Joshi, M., & Rosé, C. P. (2008). Investigating the effect of discussion forum interface affordances on patterns of conversational interactions (p. 555). ACM Press. https://doi.org/10.1145/1460563.1460650

Ward, T. B., & Kolomyts, Y. (2010). Cognition and Creativity. In *Cambridge Handbook of Creativity* (pp. 93–112).

Wu, Y., Kropczynski, J., Shih, P. C., & Carroll, J. M. (2014). Exploring the ecosystem of software developers on GitHub and other platforms (pp. 265–268). ACM Press. https://doi.org/10.1145/2556420.2556483

Yu, L., Kittur, A., & Kraut, R. E. (2014a). Distributed analogical idea generation: inventing with crowds (pp. 1245–1254). ACM Press. https://doi.org/10.1145/2556288.2557371

Yu, L., Kittur, A., & Kraut, R. E. (2014b). Searching for analogical ideas with crowds (pp. 1225–1234). ACM Press. https://doi.org/10.1145/2556288.2557378

Yu, L., Kittur, A., & Kraut, R. E. (2016). Encouraging “Outside- the- box” Thinking in Crowd Innovation Through Identifying Domains of Expertise (pp. 1212–1220). ACM Press. https://doi.org/10.1145/2818048.2820025

Yu, L., Kraut, R. E., & Kittur, A. (2016). Distributed Analogical Idea Generation with Multiple Constraints (pp. 1234–1243). ACM Press. https://doi.org/10.1145/2818048.2835201

Yu, L., & Nickerson, J. V. (2011). Cooks or cobblers?: crowd creativity through combination. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1393–1402). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1979147

Zagalsky, A., Feliciano, J., Storey, M.-A., Zhao, Y., & Wang, W. (2015). The Emergence of GitHub as a Collaborative Platform for Education (pp. 1906–1917). ACM Press. https://doi.org/10.1145/2675133.2675284

Zhang, A. X., Ackerman, M. S., & Karger, D. R. (2015). Mailing Lists: Why Are They Still Here, What’s Wrong With Them, and How Can We Fix Them? (pp. 4009–4018). ACM Press. https://doi.org/10.1145/2702123.2702194

Zhao, Z., Badam, S. K., Chandrasegaran, S., Park, D. G., Elmqvist, N. L. E., Kisselburgh, L., & Ramani, K. (2014). skWiki: a multimedia sketching system for collaborative creativity (pp. 1235–1244). ACM Press. https://doi.org/10.1145/2556288.2557394